# Negative Controls for Instrumental Variable Designs*

Oren Danieli, Daniel Nevo, Itai Walk, Bar Weinstein, Dan Zeltzer†

December 25, 2022

Click here for the latest version

**Abstract**

Instrumental variable designs rely on exclusion and independence assumptions that are not directly testable. A commonly used indirect falsification ("placebo") test for these assumptions examines whether variables that should be independent of a valid instrument are indeed independent of the instrument being used (e.g., testing that an instrument is balanced on predetermined characteristics). Adapting terminology from other disciplines, we refer to these variables as *negative controls*. We develop a theoretical framework for negative-control tests for instrumental variables and derive the assumptions underlying such tests. The theory defines negative controls as proxies for *alternative path variables* — unobserved threats to instrument validity. While many studies typically use only one negative control, our theory can be used to detect multiple negative controls in existing data sets. Our theory also shows that negative-control tests map to testing the (conditional) independence of the set of negative controls and the instrument. We discuss various conditional independence tests that can be used as negative-control tests, including tests that can also detect a violation of the functional form assumption when 2SLS is used. We demonstrate that these negative control tests can find and diagnose potential violations of the identification assumptions that commonly used tests would miss.

*We thank seminar participants at Princeton University and the Technion - Israel Institute of Technology for helpful comments and suggestions. This work was supported by a grant from Tel Aviv University Center for AI and Data Science (TAD) in collaboration with Google as part of the initiative of AI and DS for social good.

†Tel Aviv University. Danieli (corresponding author) Email: orendanieli@tauex.tau.ac.il.

# 1 Introduction

To evaluate the plausibility of the identification assumptions in instrumental variable (IV) designs, researchers often use a falsification test for the independence between the instrument and variables it should be independent of (e.g., testing balance on predetermined characteristics). Similar falsification tests are used in epidemiology (Lipsitch et al., 2010; Arnold and Ercumen, 2016; Shi et al., 2020), where researchers test for independence between a treatment and a negative-control outcome: a variable that should be independent of the treatment when the treatment is unconfounded.[1] However, formal theory for these tests in the context of instrumental variables is limited. As a result, instrumental variables falsification tests are often ad-hoc, loose about the underlying assumptions, and do not make efficient use of available data.

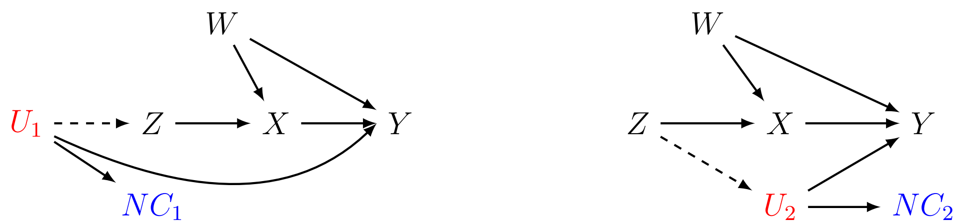
In this paper, we develop a theory of testing for instrument validity using *negative-control variables*: correlates of unobserved confounders or mediators of the instrument-outcome relationship. This theory explicates the underlying assumptions behind these common falsification tests. These assumptions can help researchers consider a broader set of negative controls. We show that under weak regularity conditions, tests for conditional independence that jointly use all available negative controls are valid falsification tests. Applying such tests to earlier studies that used instrumental variable designs, we demonstrate that they can help researchers detect and diagnose identification threats.

Figure 1 illustrates how negative controls help evaluate two main identification assumptions: independence and exclusion.[2] For concreteness, consider the context of evaluating the impact of teacher quality (by some measure $X$) on student performance (e.g., grades $Y$), with some potential confounder $W$. Therefore, an instrument $Z$ is used, such as a random or quasi-random assignment of teachers. The independence assumption implies that teacher assignment $Z$ is independent of the potential student performance (under a given assignment and teacher quality). It can be violated if, for example, student ability $U_1$, which clearly affects the grades, also affects teacher assignment. Instrument exclusion implies that the assignment affects grades only through teacher quality. It can be violated if, for example, better teachers also have better classroom amenities $U_2$ that impact student performance. We call these $U_1$ and $U_2$ *alternative-path variables* (APVs). APVs pose threats to the identification

[1] For example, epidemiologists studying the impact of flu shots on hospitalizations with respiratory conditions may check that they do not also "prevent" unrelated trauma hospitalizations (Shi et al., 2020).

[2] Throughout the paper, we use directed acyclic graphs (DAGs) as a way to visualize and communicate certain aspects of potential underlying complex structures in the discussion of negative controls. While we do not formally use DAGs, associated structural equation models, and do-calculus (Pearl, 2009) in any of our proofs, we largely interpret DAGs accordingly. We thus follow in this paper the approach advocated by Imbens (2020) that highlights DAGs as an efficient and clear exposition tool.

but are typically unobserved, so the potential violations of the identification assumptions cannot be tested directly.




(A) $NC_1 \not\perp\!\!\!\perp Z$ implies $U_1 \longrightarrow Z$, violating instrument independence.

(B) $NC_2 \not\perp\!\!\!\perp Z$ implies $Z \longrightarrow U_2$, violating instrument exclusion.

Figure 1: Negative Control Examples: Graphical Illustrations

*Notes:* The figure illustrates the use of negative control variables for evaluating the validity of instrumental-variable designs. In both panels, $X$ is the endogenous variable, $Y$ is the outcome, $Z$ is the instrument, and $W$ is a potential confounder that motivates the use of the instrument. The variables $U_1$ and $U_2$ are unobserved *alternative path variables* that pose threats to identification, in the sense that if they are related to the instrument $Z$ (through the links specified by a dashed arrow), either independence (Panel A) or the exclusion restriction (Panel B) is violated. In each case, an observed *negative control variable* ($NC_1$ or $NC_2$) that is related to the respective alternative path variable ($U_1$ or $U_2$) but not directly to $Z$ can be used to evaluate the presence of the problematic link. See Section 2 for formal definitions and results.

Negative control variables are observed proxies for underlying threats to the identification. In the example, student past grades $NC_1$ and survey results on students' satisfaction with classroom amenities $NC_2$ are two negative controls: for the instrument to be valid, they should be independent of teacher assignment $Z$. If either of these negative controls is not independent of $Z$, an alternative path must exist between the instrument and the outcome, violating the identification assumptions.

We start by clarifying the assumption underlying negative-control tests. We outline a theoretical framework that embeds negative control variables within the standard instrumental variable causal model (Angrist et al., 1996). The defining assumption that characterizes a negative control is that, if the instrument is invalid, the negative control should be related to the instrument through an APV —a variable that creates an alternative path between the instrument and the outcome. That is, a variable $NC$ is a *negative-control* if there exists an alternative path variable $U$ such that if the instrument is valid (and specifically $Z \perp\!\!\!\perp U$), then it also satisfies

$$NC \perp\!\!\!\perp Z|U.$$

Therefore, if a variable is related to a valid instrument in any way that is not through an APV (for instance, if it is directly affected by the instrument), it cannot serve as a negative control and cannot be used to test instrument validity.

We prove that if the instrument and negative control are not independent, then either independence or exclusion are violated. If a negative control is related to the instrument, some of this relation is through an APV, which implies a dependency between the APV and the instrument. Such dependency implies that there exists an alternative path between the instrument and the outcome, which violates one of the two identification assumptions. We extend this idea to designs that include instruments believed to be valid conditional on some control variables and show how these controls should be incorporated as part of the negative-control test.

Our theory implies that to test the validity of an instrument, a researcher should verify that all available negative controls are independent of the instrument (conditionally on controls). In specifications that do not require control variables, independence can be tested in various ways, such as a permutation test (similar to Ludwig et al., 2017). When the instrument is only expected to be valid conditional on some control variables, a researcher needs to perform a conditional independence test.

One such conditional independence test is the F-test, and we discuss its assumptions. The researcher can regress the instrument on all control and negative control variables and use an F-test for the null hypothesis that the coefficients of the negative controls are all zero. We show that rejection of this null could be driven by two separate problems in the identification. First, it is possible that conditional independence of the instrument and the negative controls does not hold, in which case the instrument is invalid. Alternatively, rejecting the null can also indicate a violation of “rich covariates" — a linear dependency between the instrument and the control variables that are included in the specification (Kolesár et al., 2013; Blandhol et al., 2022). In both cases, the two-stage least-squares (2SLS) analysis will result in inconsistent estimators.

More generally, conditional independence testing is an active field of research, and we discuss a few additional examples of such tests and their required assumption. Tests based on generalized additive models can also be applied to capture non-linear conditional dependencies between the instrument and the negative control. We also discuss more general tests for conditional independence that do not rely on the rich covariates assumption. These tests are suitable for cases where the instrument is not used in a 2SLS setting (e.g. Abadie, 2003).

We build on our theory to provide practical suggestions for practitioners, that can generate more powerful negative control tests. While in many cases researchers use only one negative control, we propose to use our theory to detect all negative controls that may exist in the data. We discuss common examples of negative controls and the conditions for their validity. We also discuss the practical tradeoffs in the choice of the specific testing method.

We then turn to the interpretation and diagnosis of the test outcomes in light of our theory. We illustrate that when the null hypothesis of the negative-control test is rejected, a simple diagnostic tool—comparing the correlations of negative controls with the outcome on the one hand and with the instrument on the other hand—can help highlight suspected alternative paths that might violate the validity of the instrument. When the null is not rejected, the instrument might still be invalid. Failure to reject the null could be due to insufficient power, or due to other alternative paths that are not captured by any of the observed negative controls.

We illustrate how our method works in practice by applying it to two prior works that used instrumental variable designs: Autor et al. (2013), which uses shift-share instruments to study the impact of Chinese import competition on US manufacturing employment, and Deming (2014), which uses a lottery-based instrument to evaluate the impact of school value-added on student performance. In both cases, we find that a combination of negative controls can predict the instrument, and we use our diagnostic tools to understand the nature of the violation. For Deming (2014), this diagnosis highlights a subtle endogeneity problem and can be used to amend the design and restore its validity.

The notion of negative controls originated in experimental biomedical research (e.g., Tanida et al., 2008; Power et al., 2012; Lee and Schapira, 2021). In that context, researchers often choose a *negative-control outcome*: a variable known not to be affected by the treatment but to have the same suspected confounding mechanisms as the outcome of interest. For example, a negative control can be a molecule that should not be affected by a tested drug. Then, researchers test for independence between the treatment and the negative control outcome. A rejection of such independence (i.e., when exposure to the drug is associated with a change in the molecule) is interpreted as evidence that the treatment-outcome relationship is confounded. Similar ideas are increasingly used in epidemiology (Lipsitch et al., 2010; Arnold et al., 2016; Dagan et al., 2021). This paper extends the notion of negative controls to falsification tests for instrumental variable designs and develops the underlying theory.[3]

In practice, placebo tests for instrumental variables—some of which constitute simple negative-control tests—have been frequently used in the social sciences. For example, Athey and Imbens (2017) informally discuss a common placebo test for causal inference designs: repeating the main analysis and replacing the outcome variable with a pseudo-outcome such as the lagged outcome variable. This test is consistent with our theory as pseudo-outcomes are often negative controls. Eggers et al. (2021) discuss the growing use in political science of a broader class of placebo tests, some of which correspond to negative control tests.

[3]In the context of pharmacoepidemiological studies, Davies et al. (2017) propose that a negative control approach may be used to compare instrumental variable and ordinary least square designs. But they do not formally develop such an approach.

However, the validity of negative control tests for instrumental variables is typically discussed informally and in specific contexts (e.g., Guidetti et al., 2021). Our contribution is twofold. First, we outline a general theory that uncovers these tests' underlying assumptions. This theory can help researchers systematically identify all variables in their data that can be used as negative controls. Second, our proposed approach to testing would help researchers combine information from all available negative controls instead of using them one by one in multiple separate placebo tests.

This paper proceeds as follows. Section 2 develops the theory of negative-control tests for instrumental variable designs. Section 3 discusses conditional independence testing under different assumptions. Section 4 discusses practical implementation considerations. Section 5 illustrates the proposed approach by using it to evaluate existing instrumental variable designs. Section 6 concludes.

# 2 Theory of Negative Controls in Instrumental Variable Settings

The motivation for using negative controls in tests for instrument validity is to detect potential unobserved threats to identification—specifically, to the independence and exclusion assumptions. To develop our theory, we first formalize the properties of these unobserved threats by defining *alternative-path variables* (APVs)—variables such as confounders or mediators of the instrument-outcome relationship, that could generate an alternative path between the instrument and the outcome, not through the treatment. While APVs are often unobserved, researchers can use negative control variables as observed proxies for them. We discuss the properties that negative control variables should satisfy in order to be informative for testing the instrument's validity.

## Setup

For a set of independent and identically distributed observations $i = 1, \ldots, n$, denote the observed (endogenous) treatment status by $X_i$, the candidate instrument by $Z_i$, and the observed controls by $C_i$. For simplicity of presentation, we do not include control variables in this section but we do include them in all proofs. Let $Y_i(x, z)$ be the potential outcome for unit $i$ had $X_i$ and $Z_i$ been set to the values $x$ and $z$, respectively. We make the standard assumption that the observed outcome $Y_i$ is $Y_i = Y_i(X_i, Z_i)$.

The two instrumental variable assumptions we focus on are independence and exclusion.[4]

[4] The additional assumptions of instrument monotonicity and relevance are not tested by our approach.

Independence maintains that instrument assignment is independent of the potential outcome:

**Assumption 1** (Independence). *For all $x, z$:*

$$Z_i \perp\!\!\!\perp Y_i(x, z).$$

Panel A of Figure 1 illustrates a potential violation of the independence assumption. Exclusion restriction maintains that the instrument has no direct effect on the outcome:

**Assumption 2** (Exclusion Restriction). *For all $x, z, z'$:*

$$Y_i(x, z) = Y_i(x, z') = Y_i(x).$$

Panel B of Figure 1 illustrates a potential violation of the exclusion restriction assumption. Together, independence and exclusion restriction imply the following condition:

$$Z_i \perp\!\!\!\perp Y_i(x). \tag{1}$$

Therefore, if condition (1) is not satisfied, at least one of the assumptions: independence or exclusion restriction, is violated. In loose terms, Equation (1) requires that there are no alternative paths between the instrument and the outcome except through the treatment. However, because potential outcomes are never observed, these two assumptions and the ensuing condition (1) cannot be tested directly.

## Alternative Path Variables - Single Violation

Our theory characterizes negative controls as proxies for threats to the instrument's validity. In this section, we formalize the notion of a threat by defining *Alternative Path Variables* (APVs). APVs are variables that could generate an alternative path between the instrument and the outcome. Intuitively, such variables are related to the outcome not through the endogenous treatment. Therefore, any association between them and the instrument implies that the instrument is invalid. For instance, when studying the impact of teacher quality using some quasi-random assignment of teachers, student ability or classroom amenities are APVs. If the instrument is associated with student ability or classroom amenities, it implies that the teacher allocation could be related to the outcome through its link to one of them. This corresponds to the situation in Panel A (student ability) and Panel B (classroom amenities) of Figure 1.

We start with the simple case in which only one potential threat to instrument validity exists (namely, one potential alternative path from the instrument to the outcome). In this case, an APV is any variable that satisfies the following two conditions.

**Definition 1** (Alternative-Path Variable with a Single Violation)**.** *A random variable $U$ is an APV if the following two conditions hold:*

1. *Latent instrument validity.* $Z \perp\!\!\!\perp Y(x)|U$.

2. *Path indication. If* $Z \perp\!\!\!\perp Y(x)$ *then* $Z \perp\!\!\!\perp U$.

In the next section, we provide a more general definition of an APV for cases of multiple potential threats to the identification. We show that the above definition for the single violation case is a specific case of it.

Latent instrument validity tells us that the instrument could have been valid, had only we observed and conditioned on the alternative path variable $U$. This assumption implies that $U$ represents the threat itself to the identification, as conditioning on it blocks an alternative path between the instrument and the outcome.[5] By this assumption, proxies for identification threats cannot be APVs, as controlling for a proxy does not make the instrument and the (potential) outcome independent of each other. Latent instrument validity is analogous to the latent exchangeability assumption appearing in recent literature on negative controls in epidemiology (Shi et al., 2020) and statistics (Tchetgen Tchetgen et al., 2020).

Path indication states that a valid instrument $Z$ is not linked to the alternative path variable $U$. Its contrapositive guarantees that a link between an instrument and the APV also implies a link between the instrument and the potential outcome—the link of interest for evaluating instrument validity. In the next part, we show that negative control tests search for an association between the instrument and an APV. Path indication implies that tests for association between the instrument and the APV are also tests for instrument validity.

Path indication implies that the APV is related to the outcome, and thus guarantees that the APV is informative of the instrument's validity. While in many cases the APV causally affects the outcome, there could also be APVs that are linked to the outcome in a non-causal way, as demonstrated in Example B.1. Path indication rules out variables unrelated to the outcome, which can be associated with the instrument without implying anything about its validity. For instance, consider an allocation of treatment that is done by date of birth (as in the famous Vietnam lottery used in Angrist, 1990). It is likely that events that occurred in the individual's date of birth would be correlated with lottery results. These variables (events) may satisfy latent instrument validity but they will not satisfy path indication if they are not related to the outcome of interest.

[5] Note that a single causal chain from $Z$ to $Y$ may go through several factors: $Z \to U_1 \to \ldots \to U_k \to Y(x)$. In such case, each of $U_1, \ldots, U_k$ would constitute an APV by our definition. See also Mackie (1965)

Path indication also rules out variables that could be related to both the instrument and the outcome without generating a correlation between them. One example is when $U$ is a multidimensional variable, where the dimensions associated with the outcome are not associated with the instrument.[6] We elaborate on this in Example B.2. Another example is when a variable is correlated with the outcome for some subpopulation but correlated with the instrument for a separate population. We provide such an example in section B.3.

## Alternative Path Variables - General Definition

In many cases, researchers are concerned about more than one violation. In such cases, latent instrument validity from the single violation case (Definition 1) does not hold, as controlling for $U$ does not suffice to make the instrument valid. Therefore, we turn to a more general definition of an APV, that allows for more than one potential violation of the identification. This definition introduces an additional random variable, $V$, that captures all other potential violations.

**Definition 2** (Alternative-Path Variable). *A random variable $U$ is an APV if there exists a random variable $V$ such that the following conditions hold:*

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

We discuss these defining properties one by one.

Latent instrument validity in the general case is interpreted similarly to the single-violation world. The difference is that when there is more than one threat, $U$ does not suffice to make the instrument valid, and an additional variable $V$ needs to be conditioned on. In case of multiple additional threats, $V$ could be defined as a multivariate random variable.

As in the single-violation case, path indication guarantees that $U$ represents some relevant potential violation of the instrument assumption. In the general case, this is true only conditional on $V$. Conditioning on $V$ means that special care is needed with respect to the interplay between $Z, U, V$ and $Y(x)$, as represented by the two additional properties.

Direct instrument link means that any potential link between $Z$ and $U$ is not entirely driven through $V$. The contrapositive is that if the instrument and the APV are dependent,

[6] An alternative path variable $U$ could be a multivariate if it satisfies the conditions of the definition.

then conditioning on $V$ does not suffice to eliminate this dependency. For instance, in the context of the impact of some treatment (e.g. teacher quality) on students' test scores, let $U$ be student ability. Consider the case where there is a concern that some teachers are able to avoid teaching students with disruptive behavior, which is typically correlated with low ability. In that case, $V$ cannot be defined as disruptive behavior. It is possible that conditional on disruptive behavior the instrument is independent of ability, yet still associated with ability unconditionally.

The final property, V-validity, is a more technical requirement for the variable $V$ that represents other threats. It states that a valid instrument remains valid conditional on $V$. Example B.4 demonstrates a violation of $V$-validity when $V$ is affected by both $U$ and the outcome.

Every variable that is an APV in the single violation case (Definition 1) is also an APV based on the general definition (Definition 2). Formally, let $U$ be a random variable that satisfies the two conditions of Definition 1. Taking $V$ to be some constant, we get that $U$ also satisfies Definition 2. The first two conditions are equivalent to Definition 1 and the next two conditions are trivial.

## Negative Controls for Instrumental Variables

Building on the definition of APVs, we are now ready to formalize the assumption required for a random variable $NC$ to be used as a negative control.

**Assumption 3** (Negative Control Assumption)**.** *A random variable $NC$ satisfies the negative control assumption if there exists an alternative path variable $U$ such that if $Z \perp\!\!\!\perp U$ then*

$$Z \perp\!\!\!\perp NC | U. \tag{2}$$

Therefore, *negative controls* are any variables that satisfy this assumption.

The negative-control assumption guarantees that dependency between the instrument and the negative control would be informative of the threat $U$. It rules out variables that could be associated with the instrument even if it is valid.

More formally, this assumption implies that there exists an alternative path variable $U$, such that any relation between the instrument and the negative control ($Z \not\perp\!\!\!\perp NC$) implies a relation between the instrument and $U$ ($Z \not\perp\!\!\!\perp U$). Hence, if the instrument $Z$ is associated with the negative control variable $NC$, an alternative path exists between the instrument and the outcome through $U$. Figure 1 demonstrates this idea using two examples for negative controls that satisfy this assumption.

The negative-control assumption is violated for variables that can be related to the instrument $Z$ even if the alternative path through $U$ does not exist ($Z \perp\!\!\!\perp U$). Figure 2 shows two examples of such violations. In these examples, $U_i$ (for $i = 1, 2$) is an APV. However, $N$ is not a negative control because it is related to $Z$ in other ways, making $N$ uninformative about the APV. In Panel A, the instrument and $N$ are affected by an additional factor $Q$, which is unrelated to the outcome. In Panel B, the instrument directly affects the variable $N$, violating the negative control assumption.

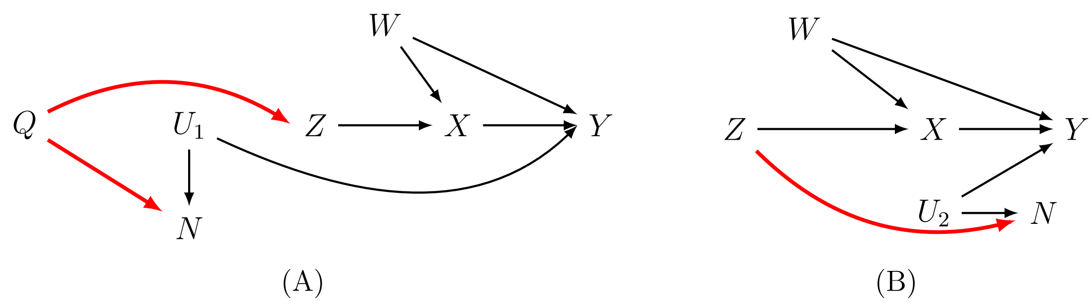



Figure 2: Violations of the Negative-Control Assumption

*Notes:* The figure illustrates two violations of the negative-control assumption. In both cases, a variable $N$ is not a valid negative control because it is related to the instrument $Z$ (through the links marked by thick red arrows), conditional on $U_j$ ($Z \not\perp\!\!\!\perp NC|U_j$ for $j = 1, 2$), while $Z \perp\!\!\!\perp U_j$. Hence, a relationship between $N$ and $Z$ is not informative about the relationship between $Z$ and $U_j$. $W$ denotes a potential confounder that motivates the use of an instrument.

The negative-control assumption implies that the treatment cannot directly affect a negative control. If a variable is affected by the treatment, it will also be affected by the instrument. Hence, for any possible APV $U$, we will have $Z \not\perp\!\!\!\perp NC|U$. For instance, Guidetti et al. (2021) criticizes the use of non-respiratory health outcomes as (what we call here) negative controls for evaluating the validity of instruments for air pollution. They argue that, because of limited healthcare infrastructure, if an instrument affects respiratory health outcomes, it may also indirectly impact non-respiratory health outcomes. In such a case, a relationship between non-respiratory health outcomes and the instrument is not informative about its validity.

The negative-control assumption is binding only for valid instruments. It allows for conditional dependency between the instrument and the negative control given the APV ($Z \not\perp\!\!\!\perp NC|U$) if the instrument is invalid ($Z \not\perp\!\!\!\perp U$). This is because some negative controls could be directly related to the instrument in cases when it is invalid. Using the previous example, consider a case where teacher allocation $Z$ is allegedly random. However, there is an allegation that the school district did not actually use randomization and instead allowed some teachers to choose their students. This could generate a link between the instrument and student ability $U$ which is an APV. The researcher can test this concern using past test scores which satisfied the negative-control assumption. In this example, if the allegations are true and the instrument is invalid it is possible that some teachers used lagged test scores directly to select students for their classes. In this case, we would get $Z \perp\!\!\!\perp NC|U$ only if the instrument is valid.

A negative-control test is any statistical test for dependence between the instrument and a negative control variable. The null hypothesis is $H_0 : Z \perp\!\!\!\perp NC$. The following theorem formalizes the implication of rejecting this null for the instrument validity.

**Theorem 1** (Negative-Controls Test)**.** *Assume that $NC$ is a negative-control variable. If $Z \not\perp\!\!\!\perp NC$, then either the independence assumption or the exclusion restriction are violated, so $Z$ is not a valid instrument.*

In Appendix A, we prove a more general version of this theorem that includes control variables. The idea behind the proof is intuitive. If the negative-control test failed (i.e., $Z \not\perp\!\!\!\perp NC$) then, because of the negative-control assumption, any dependency between the instrument and the negative control must also operate through some alternative path variable $U$. Therefore, the instrument is also related to the APV ($Z \not\perp\!\!\!\perp U$). By the definition of APV, this implies an alternative path between the instrument and the outcome, such that $Z \not\perp\!\!\!\perp Y(x)$. This can only occur if the instrument violates either the exclusion or the independence assumptions.

While the theorem shows that failing a negative-control test implies that the instrument is invalid, the opposite is not necessarily true. An invalid instrument could still pass a negative-control test. For instance, there might exist an additional APV that is correlated with the instrument, for which no negative controls are available. This concern is in addition to the regular reasons for why statistical tests can only reject and never fully prove null hypotheses. We further discuss this point in Section 4.4.

## Observed APVs

Negative controls are observed proxies for potentially unobserved threats to the identification (APVs). If an APV is observed by the researcher, it can be used directly for testing:

**Proposition 1.** *An APV satisfies the negative-control assumption (Assumption 3).*

The proof is immediate by recognizing that for every alternative path variable $U$, it is trivial that $Z \perp\!\!\!\perp U|U$. Therefore $U$ is also a negative control based on the negative-control assumption. From Theorem 1, any correlation $Z \not\perp\!\!\!\perp U$ implies that the instrument is invalid.

However, the reverse argument of Proposition 1 is not true, as a negative control is not always an APV. Intuitively, the key difference between a negative control and an APV is that negative controls can also be a proxy for a threat, and not the threat itself. In the single violation case, negative controls do not necessarily satisfy latent instrument validity. In the general case, there is no choice of a variable $V$ that satisfies all four conditions.[7] Consequently, controlling for a negative control (which is not an APV) does not eliminate the threat to instrument validity.

In some cases, an invalid instrument could become valid conditionally on an observed APV. A potential instrument might fail the negative-control test and be correlated with an observed APV. Since the APV itself is observed, the researcher can condition on it, which could potentially make the instrument conditionally valid. Clearly, whether an APV (or any other variable) should be included as a control variable depends on the specific domain of the problem. In some cases, the inclusion of control variables can result in new violations of the identification assumptions (Pearl, 2009).

## Multiple Negative Controls

The definitions and theoretical results we have presented so far apply to a multivariate negative-control vector, as long as this vector satisfies the negative-control assumption.

Moreover, as we will discuss in the next sections, multiple variables individually satisfying the negative-control assumption exist in many datasets. Let $\overline{NC} = \{NC_1, ..., NC_M\}$ be a vector of $M$ negative controls, each satisfying the negative-control assumption. In such cases, can the researcher test jointly whether $Z \perp\!\!\!\perp \overline{NC}$ instead of testing for independence one-by-one? In practice, the answer is typically yes. But theoretically, there are knife-edge cases in which a valid instrument which is uncorrelated with any of the individual negative control variables might be correlated with the vector $\overline{NC}$.

To test multiple negative control jointly, the researcher needs to assume such knife-edge cases do not exist. Specifically, the researcher need to assume that if $Z \perp\!\!\!\perp NC_i$ for all $i$, then $Z \perp\!\!\!\perp \overline{NC}$. Under this assumption, if the instrument is dependent with the vector $\overline{NC}$, it is also dependent with at least one of the individual negative controls, and therefore the instrument is invalid.

This assumption rules out cases where even though there exists some function of the negative controls that satisfies $Z \not\perp\!\!\!\perp g(NC_1, ..., NC_M)$ the instrument is independent with every negative control individually ($Z \perp\!\!\!\perp NC_i$ for all $i$). This can only occur for very particular data

[7] A negative control $NC$ could satisfy latent instrument validity for a variable $V$ which is the relevant APV for this negative control ($Z \perp\!\!\!\perp NC|V$). However, for this choice of $V$, the negative control will not satisfy direct instrument link.

generating processes. Small deviations from such data generating process would typically generate a dependence with one of the negative controls.[8] Example B.5 provides such a case using the XOR operator. We also demonstrate how small changes to this data-generating process would generate dependence with one of the negative controls.

### Negative Control Exposure

Previous papers on negative controls in other disciplines distinguish between negative control outcome and negative control exposure (NCE) (Shi et al., 2020; Tchetgen Tchetgen et al., 2020). So far we discussed the more common type of negative-control tests which examine the link between the negative control and the instrument. This is similar to the concept of negative control outcomes, where researchers test for dependencies between the treatment and the negative control. In some cases, which are less frequent in applied economics, researchers use a negative-control test to examine a link between the negative control and the outcome. This corresponds to the case of NCE.

For instance, Arteaga and Barone (2022) studies the causal effect of marketing efforts on opioid addiction. To instrument for marketing efforts, they use local cancer mortality, as opioids were initially marketed for cancer patients. To test the validity of the instrument, Arteaga and Barone (2022) use local mortality from other causes as NCEs. They then show that these NCEs are independent of the outcome (opioid addiction). This is different from the more standard way of using the negative-control test, which is to test for dependency with the instrument.

Formalizing NCEs requires modifications to the APV definition and the negative control assumption, which we plan to include in the next version of this paper. Appendix A.2 shows typical examples for usage of NCE for instrumental variables. The rest of the paper focuses on the more widespread use of negative control outcomes, where researchers test for links between the negative control and the instrument.

## 3 Falsification Tests Using Negative Controls

Theorem 1 motivates a general approach for searching potential violations of the independence or exclusion assumptions using negative controls. In cases without control variables, if the null hypothesis $H_0 : Z \perp\!\!\!\perp NC$ is rejected based on the data, one can conclude that the instrument is not valid. This hypothesis can be tested via a variety of independence

[8]In the DAGs literature, the exclusion of small number of probability functions that generates these unlikely independence structures is formalized via the notion of *stability* or *faithfulness*; see (Pearl, 2009, Page 48).

tests (e.g., Székely et al., 2007; Heller et al., 2013). For instance, one could try to predict the instrument using the negative controls, and compare this to predictions of permutated versions of the instrument for inference (similar to Ludwig et al., 2017).

In the common and more general case when controls are included in the analysis, researchers should test the null hypothesis of conditional independence $H_0 : Z \perp\!\!\!\perp NC|C$. Appendix A presents the definitions of independence and exclusion, and develops our theory for negative-control testing for specifications that include control variables.[9] Conditional independence testing arises in numerous practical examples, among which causal inference, graphical models, and detecting genetic associations. As a result, a plethora of methods to test conditional independence exist, and new methods are still frequently introduced as this is an active field of research. A key recent result is that non-parametric testing for conditional independence when $C$ includes continuous variables is a hard statistical task for which there is no single optimal test procedure (Shah and Peters, 2020). Shah and Peters deduced from their result that some domain knowledge should be used to select an appropriate statistical test.

With this in mind, we start by discussing conditional independence tests in 2SLS specification. 2SLS specifications rely on an additional functional form assumption that we can use in our testing. We then discuss more general non-parametric conditional independence tests. For a more general survey of the methods discussed in this section and other related methods in the context of causal discovery see Heinze-Deml et al. (2018).

## 3.1 Tests for Conditional Independence in 2SLS Specifications

While instruments can be used in a variety of causal methods (e.g. Abadie, 2003), in the large majority of applications, they are used in a 2SLS specification. The 2SLS analysis is only valid under the following additional assumption (Blandhol et al., 2022):

**Assumption 4** (Rich Covariates)**.** *The instrument $Z$ is linear in the controls. Namely:*

$$E[Z|C] = \gamma C, \tag{3}$$

*where $C$ is the vector of control variables used in a 2SLS specification.*

Therefore, for an instrument $Z$ satisfying the independence, exclusion restriction, and rich covariates assumptions and for a valid negative control $NC$, we will expect that

$$E[Z|C, NC] = \gamma C,$$

[9] $NC$ and $C$ may be, and often are, multivariate.

which provides a more specific null hypothesis for conditional independence testing. That is, we can write the conditional expectation as

$$E[Z|C, NC] = g(C, NC), \tag{4}$$

for some function $g$. We can then test the null hypothesis:

$$H_0 : g(C, NC) = \gamma C. \tag{5}$$

If this null hypothesis is rejected, it implies that at least one of the following assumptions is not satisfied: independence, exclusion, or rich covariates. In either case, the 2SLS analysis is invalid. If researchers are willing to assume that the rich covariates assumption holds (e.g., when the model is fully saturated in the controls), a rejection of the null hypothesis (5) implies that either independence or exclusion is violated.

With additional assumptions, this test boils down to the F-test that is commonly used in practice. Specifically, in addition to the negative-control assumption, the F-test requires two additional parametric assumptions. First, it postulates that $g$ is a linear function, and hence $Z$ can be written using the following linear model

$$Z = \beta_0 + \beta_C^T C + \beta_{NC}^T NC + \epsilon_Z, \tag{6}$$

where $E(\epsilon_Z|C, NC) = 0$. Conditional independence between $Z$ and $NC$ given $C$ boils down to $\beta_{NC}^T$ equaling zero. If in addition, the errors are normally distributed $\epsilon_Z \sim N(0, \sigma_\epsilon^2)$, then one could use the standard F-test for the null hypothesis $H_0 : \beta_{NC}^T = 0^T$, where $0^T$ is a vector of zeros.

The assumption of linear dependence between the instrument and the negative control can be relaxed. Consider, for example, the use of General Additive Models (GAMs) (Wood, 2006). One such specification assumes that:

$$Z = g(C, NC) + \epsilon_Z = \sum_{j=1}^{J} g_j^{(C)}(C_j) + \sum_{k=1}^{K} g_k^{(NC)}(NC_k) + \epsilon_Z, \tag{7}$$

where $J$ and $K$ are the number of controls and negative controls, respectively, $g_j^{(C)}$ and $g_k^{(NC)}$ are smooth functions, and $\epsilon_Z$ are independent errors with $E(\epsilon_Z|C, NC) = 0$. Under this model, $g(C, NC)$ is assumed to be composed of additive smooth functions that are typically estimated using splines (Wood, 2006). When rich covariates is assumed $g_j^C$ is restricted to be linear. Interactions can be accommodated by including additional smooth functions of $C_j \cdot C_{j'}$, $C_j \cdot NC_k$, or $NC_k \cdot NC_{k'}$ for some $j, j', k$ or $k'$. Under normal errors assumption, an approximate generalized likelihood ratio test can be constructed to test the null hypothesis (Equation (5)) versus the alternative that $g(C, NC)$ can be written more generally as in Equation (7).[10]

### 3.2 Nonparametric Conditional Independence Tests

If researchers are not using 2SLS and are not willing to assume rich covariates (Assumption 4), more general types of nonparametric conditional independence tests exist. One approach, which is more similar to common practices in economics, is invariant target prediction (Heinze-Deml et al., 2018). This method suggests using a prediction algorithm to predict the instrument twice, once using only the controls and once using both the negative controls and the controls. Under $H_0$, the out-of-sample performance of both predictions would be close to identical. While many prediction algorithms—including those based on parametric models—can be used, this method typically uses more flexible, non-parametric models or algorithms.[11]

While invariant target prediction typically focuses on mean-independence, a more general approach uses kernel measures. These tests rely on the fact that, under certain conditions, conditional independence is equivalent to a conditional cross-covariance operator defined for reproducing kernel Hilbert spaces equaling zero (Fukumizu et al., 2007; Zhang et al., 2011; Strobl et al., 2019).[12] As a result, a test for zero Hilbert-Schmidt norm of the conditional cross-covariance operator is a test for conditional independence. Zhang et al. (2011) derived the asymptotic distribution of an estimator for this norm under the null and proposed a simulation-based method to calculate the null distribution and a Gamma distribution approximation.

## 4 Practical Considerations

In this section, we discuss practical considerations for researchers that want to use a negative-control test for their instrument validity. We recommend that researchers follow two steps

[10]Technical details of how to construct such a test and conditions for its validity can be found in Sections 6.12.3 and 6.12.4 in Wood (2006).

[11]This is similar to the multiple-hypotheses test suggested by Ludwig et al. (2017) when control variables are included. For more details, see Appendix B.4 in Heinze-Deml et al. (2018).

[12]For unconditional independence, the cross-covariance operator $\Sigma_{Z,NC}$ is the covariance between $h_1(Z)$ and $h_2(NC)$ for all $h_1$ and $h_2$ in the reproducing kernel Hilbert spaces defined on the domains of $Z$ and $NC$, respectively. The conditional cross-covariance operator conditionally on $C$ is then defined by $\Sigma_{Z,NC} - \Sigma_{Z,C}\Sigma_{C,C}^{-1}\Sigma_{C,NC}$. However, the relevant conditional cross-covariance operator is the one defined between $(Z, C)$ and $NC$ conditionally on $C$.

to evaluate the validity of instrumental variable designs. First, using domain knowledge in their empirical setting, researchers should select all variables that satisfy the negative-control assumption. We discuss this selection process in Section 4.1. Second, researchers should select the appropriate statistical test, which depends on the size of the sample, the number of negative controls (and other control variables), and functional form assumptions. We discuss this step in Section 4.2. The common practices for the usage of negative controls in economic research in light of these recommendations are discussed in Section 4.3. The aftermath of the test is discussed in section 4.4. We suggest diagnostic methods for cases when the test rejects the null and discuss the interpretation of not rejecting it. We demonstrate these principles using two examples in Section 5.

## 4.1 Choosing Negative Controls

We recommend that researchers would try to identify variables that satisfy the negative-control assumption that exist or could be constructed in their data sets. As we illustrate in the examples below, such variables that can easily be extracted from the data are sometimes left unused. We now discuss some common types of variables that are used in falsification tests and the assumptions underlying their validity as negative controls.

**Predetermined events.** Predetermined variables can often serve as useful negative controls. For example, parental education may serve as a negative control in designs that are based on a child's quarter of birth. But not every predetermined variable is a good negative control. In particular, predetermined events can sometimes have a direct causal effect on the instrument, which would violate the negative-control assumption. For example, parents' quarter of marriage is not a valid negative control for the instrument quarter of birth as it is likely affecting it. Predetermined events would always satisfy the negative-control assumption for instruments that are based on randomization (e.g., school lotteries).[13]

**Repeated luck.** Some instruments are based on certain occurrences being akin to a strike of luck. The validity of such instruments is questioned when this strike of luck systematically repeats itself. For instance, Jäger and Heining (2019) uses a worker's premature death as an instrument for worker turnover, and argues that such deaths occur randomly across firms. To support this claim, they show that workers' premature death does not predict additional premature deaths within the same firm. If such a correlation existed, this would suggest that firms with premature deaths are potentially riskier, which could affect outcomes directly, not

[13] In this context, by randomization we mean that the distribution for the instrument $P(z|c)$ will not change with any additional information $P(z|c) = P(z|c, \mathcal{I})$.

through turnover. The negative-control assumption requires that the instrument will not have any causal effect on its future measurements.

**Unaffected outcomes.** Another set of potential negative controls includes alternative outcomes that we do not expect to be affected by the instrument. For instance, imagine one tries to study the impact of a reading tutoring program on reading scores, using some quasi-random allocation of this tutoring. One may choose to test that the quasi-random allocation has no effect on alternative seemingly unrelated outcomes, such as math scores. Such a relation could occur if the allocation is not quasi-random (violation of independence), or if the allocation is bundled with other resources that the student is getting (violation of exclusion). Alternative outcomes require a careful examination of the negative-control assumption using domain knowledge. The assumption would be violated if the treatment ($X$) has a direct effect on the alternative outcome. For instance, additional reading tutoring could come at the expense of time dedicated to studying math, which would affect math scores. The assumption would also be violated if the main outcome ($Y$) affects the alternative outcomes. For instance, improving reading skills could help students interpret math problems.

**What not to include.** Negative controls without a plausible relationship to any APV should never be included. Adding negative controls with no or very weak relationship with an APV can reduce the power of the test. This is analogous to adding spurious controls to an ordinary least squares regression. Hence, researchers may want to exclude negative controls with only a tenuous link to potential APVs.

## 4.2 Choosing a Statistical Test

When using 2SLS with controls (which is the common use of 2SLS, Blandhol et al., 2022), researchers rely on the instrumental variable assumptions (independence and exclusion), as well as on the rich covariates assumption (4). Therefore, we recommend implementing one of the methods in Section 3.1 that tests for the three assumptions jointly. These methods focus on whether negative controls improve the prediction of the instrument compared to a prediction that uses only control variables. This follows the advice of Shah and Peters (2020) to use domain knowledge (rich covariates) for selecting the appropriate conditional independence test.

The choice of method should depend on sample size limitations and contextual knowledge of the importance of non-linearity in the relationship between a negative controls and the instrument. In smaller samples, or in cases where linearity is a reasonable approximation, an F-test is expected to perform better. In the next version of the paper, we will present simulation results demonstrating these trade-offs.

## 4.3 Discussion of Common Practices

A common practice for negative control is using them as "pseudo outcomes" (Athey and Imbens, 2017). The researcher replaces the outcome variable in the reduced form specification with a negative control. This practice is different from the methods we discussed in Section 3.1 that discussed using the instrument as the outcome variable. Similar to our discussion in Section 3.1, this is a valid negative-control test, under the rich covariates assumption. Hence, a rejection of the test could imply either a problem with the instrument's validity or with the linear specification.

Such a test using a single negative control is typically less powerful than using multiple negative controls, which, as previously discussed, often exist in many data sets. In many cases, researchers do not use all the available negative controls in their data, which leads to a loss of information. When multiple negative controls are used, this method creates a challenge of multiple hypotheses testing. Various methods for multiple hypotheses testing can be applied (e.g., Bonferroni). However, these methods do not aggregate information efficiently (Ludwig et al., 2017). Furthermore, these methods tend to be conservative with respect to the type I error control, leading to limited power to unravel potential violations of Independence or Exclusion. The reverse problem of predicting the instrument with the negative control allows researchers various methods to aggregate information into a single hypothesis, as we discussed.

## 4.4 Interpreting the Test Results

### Diagnosing failures

In this section, we will suggest tools that can help researchers understand the reason why a negative control test failed. If one of the tests from Section 3.1 were used, a failure of the test (rejection of the null) could imply a violation of either the instrument assumptions or the rich covariates assumption. In this case, we suggest that researchers would test the rich covariates directly. For instance, this can be done using a Ramsey RESET test using only the control variables, without negative controls (Ramsey, 1969). Conversely, researchers can also perform one of the tests in Section 3.2 that do not depend on the rich covariates assumption, and test the instrument validity directly.

Assuming that the rich covariate assumption holds, a failure of the negative-control test implies that there exists an alternative path between the instrument and the outcome. When a test using multiple negative controls fails, researchers might be interested in evaluating the relative contribution of different alternative paths to the failure. Since in many cases researchers cannot observe the APV themselves, they can use the negative controls that proxy for them. Figure 3 demonstrates this for a case of an unobserved confounder. The researcher wants to learn about the link between the instrument $Z$ and the outcome $Y$, which operates through the alternative path variable $U$. This link is marked with arrows A and B. However, since $U$ is unobserved, the researcher can only measure the link between $Z$ and $NC$ (arrows A and C) and between $Y$ and $NC$ (arrows B and C). This can be done by considering which negative controls are most associated with the instrument on the one hand and with the outcome on the other hand. For instance, by using correlations between the negative controls and the instrument or outcome, or by using random forest and variable importance measures when predicting the instrument or outcome with the negative controls. We now discuss the interpretation of these associations, as well as their limitations.

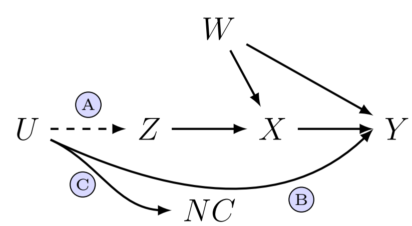



Figure 3: Diagnosing Failures

*Notes:* A potential threat operates through the alternative path variable $U$. The researcher is particularly interested in the strengths of the relations between $U$ and the instrument (Arrow A) and between $U$ and the outcome (Arrow B). However, in general, $U$ is not directly observed. Instead, the researcher can proxy for these links using the negative controls, and consider which negative controls are most predictive of the instrument on the one hand and of the outcome on the other hand. However, these tests are also influenced by the relationship between the negative control and the APV (Arrow C). $W$ denotes a potential confounder that motivates the use of an instrument.

First, examining which negative controls are highly predictive of the instrument could help the researcher understand the reasons for the violation of the assumptions. This helps the researcher to understand which negative controls lead the negative control test to fail. If the researcher knows which APV each negative control is proxying for, they could understand which alternative paths exist. In Figure 3, this association corresponds to the arrows A and C. Such an association implies that $U$ and $Z$ are linked (Arrow A) and so the instrument is invalid.

Second, examining which negative controls are highly predictive of the outcome could help the researcher understand which violations are more substantial. The threat to identification is larger the stronger the link between the APV and the outcome. Alternative paths that are more directly related to the outcome, would generate a larger bias in the results. Since the APV is not observed, we recommend the researcher to measure the association between the outcome and the negative control. In Figure 3, this association corresponds to arrows B and C. This exercise can help researchers understand the magnitude of the potential biases that different APVs are posing (Arrow B).

For example, a researcher who uses quasi-random teacher assignments may be concerned that the allocation was somehow influenced by each of two (unobserved) alternative paths from the instrument to the outcome: student ability or student neighborhood. To proxy for these threats, the researcher can use lagged test scores and parents' job location as negative controls, respectively. If the negative-control test fails, learning which negative control is driving the association with the instrument could help the researcher understand the nature of the violation of the identification assumptions. Further, a greater correlation between each of these negative controls and the outcome would increase the concern that the alternative path associated with it is substantial.

Overall, negative controls that are strongly correlated with both the instrument and the outcome suggest that there exists a particularly strong alternative path between the instrument and the outcome, and hence a more substantial violation of the instrument assumptions. However, the opposite is not true. Weak relations to the instrument or the outcome could be the result of a weak link between the negative control and the APV. This is marked with arrow C in our example on Figure 3. Therefore, while this analysis is useful in finding important violations of the identification assumption, it cannot rule out that other important violations exist as well. Practically, we suggest plotting the strengths of these two associations for each negative control using a two-dimensional scatter plot. We will demonstrate this Section 5.

### Interpreting Successes

When the null hypothesis of a negative-control test cannot be rejected, this does not imply that the instrument is necessarily valid. There are at least two different reasons why the null of a negative-control test might not be rejected even if the instrument is invalid. First, the instrument assumption could be violated by different alternative paths than the tested ones. For instance, imagine a quasi-random allocation to teachers that is uncorrelated with students' neighborhoods. The allocation could still be associated with students' abilities within neighborhoods, thus violating the instrument assumption.

Second, invalid instruments could pass the test due to insufficient power. In negative-control tests, the roles of type I and type II errors are replaced. Namely, for all tests described in Section 3, an $\alpha$-level test for conditional independence, will reject the null at probability

$\alpha$ at most if the instrument is valid. If the instrument is invalid, however, the ability of the data to reveal it is limited and depends on the power of the test, which is not controlled for by the researcher.[14]

# 5 Applied Examples

To illustrate how to use our method in practice, we apply it to popular instrumental variable designs used in prior work.

### Example I: Shift-Share Instruments

In their seminal work, Autor et al. (2013) (henceforth ADH) analyze the impact of Chinese import competition between 1990–2007 on US local labor markets. To address the endogeneity of exposure to imports in each US commuting zone, ADH use a shift-share instrument based on the industry exposure to Chinese imports in other developed countries. We apply negative-control tests to this instrument using data from the study replication data. Specifically, we attempt to predict the (year 2000) instrument, conditional on the same sets of original controls, using as negative controls all pre-determined economic characteristics of the commuting zone from 1970–1990. The assumption is that any relationship between the instrument and these pre-determined characteristics is driven by a characteristic of the commuting zone, which also affects manufacturing employment in 1990-2007. We use two alternative testing specifications: F-test, and Bonferroni-adjusted t-tests. The first column of Table 1 summarizes the details of our negative control test. Appendix C.1 discusses the implementation details.

Most of our negative-control tests reject the null hypothesis: US commuting zones that experienced more exposure to Chinese imports are predictably different using negative controls. Table 2 displays the test p-values using different combinations of control variables that were used in the original paper. Using the F-test, negative controls improve the prediction of the instrument under all combinations of controls used in the original study, including when the full set of original controls is included. These results suggest that areas exposed to import competition may have had different manufacturing employment trends.

These findings are perhaps not surprising, given that in their paper, ADH also found that some falsification tests have failed. ADH regress changes in manufacturing employment share in previous decades on the instrument. One of their tests showed a correlation between

[14]See Bilinski and Hatfield (2018) for a discussion of this issue and an alternative approach.

the instrument and lagged outcomes (suggesting that the estimated effect of the China shock is even larger).

In contrast to the original paper, our test is able to reject the null, even in a specification that includes all control variables. ADH used fewer lagged variables and evaluate the correlation of each one of them with the instrument separately. We repeat a similar exercise and use only lagged outcome as the negative control, using the same control variables. The third row of Table 2 presents the p-value from this exercise. We find that with control variables we cannot reject the null. This implementation demonstrates how, based on the theory of negative controls, one can develop more extensive tests that more systematically use all information available in the data jointly to evaluate the validity of an instrument.

To examine which particular negative controls raise concerns about the instrument, we use the method discussed in Section 4.4. Figure 4 shows the absolute correlation with the outcome and the instrument for each negative control. The high correlation of variables on the top right (such as the lagged share of manufacturing employment of different subgroups: workers aged 34–49, 50–64, and female workers) with both the instrument and the outcome suggests that they are related to both. Therefore these negative controls present the strongest evidence for potential threat to the instrument assumptions. This suggests that a rise in Chinese import competition is related to a different composition of workers in the manufacturing sector.

## Example II: School-Choice Lotteries

Deming (2014) evaluates whether school value-added measures capture the causal effect on student performance. To deal with selective sorting across schools, Deming exploits school admission lotteries in oversubscribed schools to construct his instrument: he considers the value-added of the student's first choice school in case they won the lottery and the value-added of the default neighborhood school in case they lost. As custom in this type of analysis, Deming uses lottery fixed effects, to control for the different probabilities of winning the lotteries for students with different choice sets.

We apply negative-control tests to this instrument using data from the study replication data. Specifically, we attempt to predict the instrument, conditional on the same sets of original controls, using past test scores and the value added of the different schools students submitted as their preferences as negative controls. The second column of Table 1 summarizes the details of our negative control test. Appendix C.1 discusses the implementation details.

Even though true randomness is involved, both the F-test and the Bonferroni approach reject the conditional independence between the instrument and the negative controls. Column 1 of Table 3) shows the p-values from both approaches. We find that both methods reject the null, and find a link between the negative controls and the instrument.

To examine which particular negative controls raise concerns about the instrument, we once again use the method discussed in Section 4.4. Panel A of Figure 5 shows, for each negative control, its absolute correlation with the outcome and the instrument. The figure reveals that the neighborhood school value-added is clearly correlated with the instrument, even conditional on the set of controls.

In this case, the problem with the instrument is solvable. Inspecting the regression specifications in Deming (2014) shows that this is a problem of insufficient controls: while the inclusion of school-lottery fixed-effects controls for differences in students' choices, it does not control for differences in their neighborhood school.[15] However, this issue can be remedied by using the purely random part of the instrument—the lottery assignment. Panel B shows that in this case, as expected, the correlation between all negative controls and the (modified) instrument is negligible. Column 2 of Table 3 confirms that this instrument indeed passes the F-test for conditional independence ($p = 0.648$).

# 6 Conclusion

This paper develops a theoretical framework for negative-control tests for instrumental variables. The theory shows that negative control tests map to testing the conditional independence of the set of negative controls and the instrument. These tests are valid if the negative controls satisfy the negative-control assumption. This assumption requires that any link between the instrument and the negative control is operating at least partially through an APV: a variable that if linked to the instrument, creates an alternative path between the instrument and the outcome.

We build on our theory to discuss a set of tests for negative controls. The theory can be used to help researchers detect several negative controls in their data set. We suggest several non-parametric conditional independence tests that can be used as negative control tests when multiple negative controls are detected. We also propose parametric negative control tests, which not only examine the instrument validity but also the validity of the 2SLS specification. We provide the reader with a practitioner's guide for designing their own negative control test. Finally, we demonstrate how these principles can be used in practice

[15] Since there is a likely alternative path between neighborhood school value-added and student test score, including it in the instrument requires controlling for its variation. This could be done for instance by including neighborhood fixed effects as well.

to detect identification problems using two different applications.

# References


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

**Angrist, Joshua D**, "Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records," *The american economic review*, 1990, pp. 313–336.

**_ , Guido W Imbens, and Donald B Rubin**, "Identification of causal effects using instrumental variables," *Journal of the American statistical Association*, 1996, *91* (434), 444–455.

**Arnold, Benjamin F and Ayse Ercumen**, "Negative control outcomes: a tool to detect bias in randomized trials," *JAMA*, 2016, *316* (24), 2597–2598.

**_ , _ , Jade Benjamin-Chung, and John M Colford Jr**, "Brief report: negative controls to detect selection bias and measurement bias in epidemiologic studies," *Epidemiology (Cambridge, Mass.)*, 2016, *27* (5), 637.

**Arteaga, Carolina and Victoria Barone**, "A Manufactured Tragedy: The Origins and Deep Ripples of the Opioid Epidemic," Technical Report, Working paper. http://www. carolinaarteaga. com/s ... 2022.

**Athey, Susan and Guido W Imbens**, "The state of applied econometrics: Causality and policy evaluation," *Journal of Economic Perspectives*, 2017, *31* (2), 3–32.

**Autor, David H, David Dorn, and Gordon H Hanson**, "The China syndrome: Local labor market effects of import competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–68.

**Bilinski, Alyssa and Laura A Hatfield**, "Nothing to see here? Non-inferiority approaches to parallel trends and other model assumptions," *arXiv preprint arXiv:1805.03273*, 2018.

**Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky**, "When is TSLS actually LATE?," Technical Report, National Bureau of Economic Research 2022.

**Dagan, Noa, Noam Barda, Eldad Kepten, Oren Miron, Shay Perchik, Mark A Katz, Miguel A Hernán, Marc Lipsitch, Ben Reis, and Ran D Balicer**, "BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting," *New England Journal of Medicine*, 2021.

**Davies, Neil M, Kyla H Thomas, Amy E Taylor, Gemma MJ Taylor, Richard M Martin, Marcus R Munafò, and Frank Windmeijer**, "How to compare instrumental variable and conventional regression analyses using negative controls and bias plots," *International Journal of Epidemiology*, 2017, *46* (6), 2067–2077.

**Deming, David J**, "Using school choice lotteries to test measures of school effectiveness," *American Economic Review*, 2014, *104* (5), 406–11.

**Eggers, Andrew C, Guadalupe Tuñón, and Allan Dafoe**, "Placebo tests for causal inference," *Unpublished manuscript. https://pelg. ucsd. edu/Eggers_2021. pdf*, 2021.

**Fukumizu, Kenji, Arthur Gretton, Xiaohai Sun, and Bernhard Schölkopf**, "Kernel measures of conditional dependence," *Advances in neural information processing systems*, 2007, *20.*

**Guidetti, Bruna, Paula Pereda, and Edson Severnini**, ""Placebo Tests" for the Impacts of Air Pollution on Health: The Challenge of Limited Health Care Infrastructure," in "AEA Papers and Proceedings," Vol. 111 2021, pp. 371–75.

**Heinze-Deml, Christina, Jonas Peters, and Nicolai Meinshausen**, "Invariant causal prediction for nonlinear models," *Journal of Causal Inference*, 2018, *6* (2).

**Heller, Ruth, Yair Heller, and Malka Gorfine**, "A consistent multivariate test of association based on ranks of distances," *Biometrika*, 2013, *100* (2), 503–510.

**Imbens, Guido W**, "Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics," *Journal of Economic Literature*, 2020, *58* (4), 1129–79.

**Jäger, S and J Heining**, "How substitutable are workers? evidence from worker deaths. Evidence from Worker Deaths," *Working Paper*, 2019.

**Kolesár, Michal et al.**, "Estimation in an instrumental variables model with treatment effect heterogeneity," Technical Report 2013.

**Lee, Jinyoung and Matthieu Schapira**, "The promise and peril of chemical probe negative controls," *ACS Chemical Biology*, 2021, *16* (4), 579–585.

**Lipsitch, Marc, Eric Tchetgen Tchetgen, and Ted Cohen**, "Negative controls: a tool for detecting confounding and bias in observational studies," *Epidemiology (Cambridge, Mass.)*, 2010, *21* (3), 383.

**Ludwig, Jens, Sendhil Mullainathan, and Jann Spiess**, "Machine-learning tests for effects on multiple outcomes," *arXiv preprint arXiv:1707.01473*, 2017.

**Mackie, John L**, "Causes and conditions," *American philosophical quarterly*, 1965, *2* (4), 245–264.

**Pearl, Judea**, *Causality*, Cambridge university press, 2009.

**Power, Jonathan D, Kelly A Barnes, Abraham Z Snyder, Bradley L Schlaggar, and Steven E Petersen**, "Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion," *Neuroimage*, 2012, *59* (3), 2142–2154.

**Ramsey, James Bernard**, "Tests for specification errors in classical linear least-squares regression analysis," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1969, *31* (2), 350–371.

**Shah, Rajen D. and Jonas Peters**, "The hardness of conditional independence testing and the generalised covariance measure," *The Annals of Statistics*, 2020, *48* (3), 1514 – 1538.

**Shi, Xu, Wang Miao, and Eric Tchetgen Tchetgen**, "A selective review of negative control methods in epidemiology," *Current Epidemiology Reports*, 2020, *7* (4), 190–202.

**Strobl, Eric V., Kun Zhang, and Shyam Visweswaran**, "Approximate Kernel-Based Conditional Independence Tests for Fast Non-Parametric Causal Discovery," *Journal of Causal Inference*, 2019, *7* (1), 20180017.

**Székely, Gábor J, Maria L Rizzo, and Nail K Bakirov**, "Measuring and testing dependence by correlation of distances," *The Annals of Statistics*, 2007, *35* (6), 2769–2794.

**Tanida, Isei, Toshiyuki Yamaji, Takashi Ueno, Shoichi Ishiura, Eiki Kominami, and Kentaro Hanada**, "Consideration about negative controls for LC3 and expression vectors for four colored fluorescent protein-LC3 negative controls," *Autophagy*, 2008, *4* (1), 131–134.

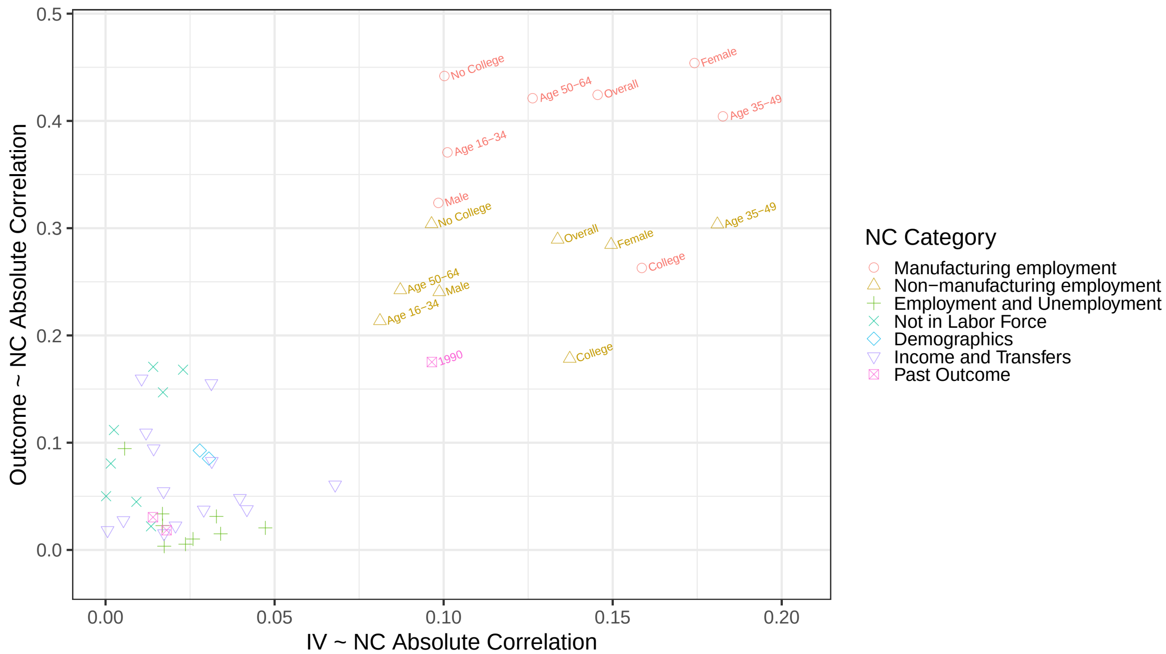
**Tchetgen, Eric J Tchetgen, Andrew Ying, Yifan Cui, Xu Shi, and Wang Miao**, "An introduction to proximal causal learning," *arXiv preprint arXiv:2009.10982*, 2020.

**Wood, Simon N**, *Generalized additive models: an introduction with R*, Chapman and Hall/CRC, 2006.

**Zhang, Kun, Jonas Peters, Dominik Janzing, and Bernhard Schölkopf**, "Kernel-based conditional independence test and application in causal discovery," in "Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence" 2011, pp. 804–813.

# Additional Figures and Tables

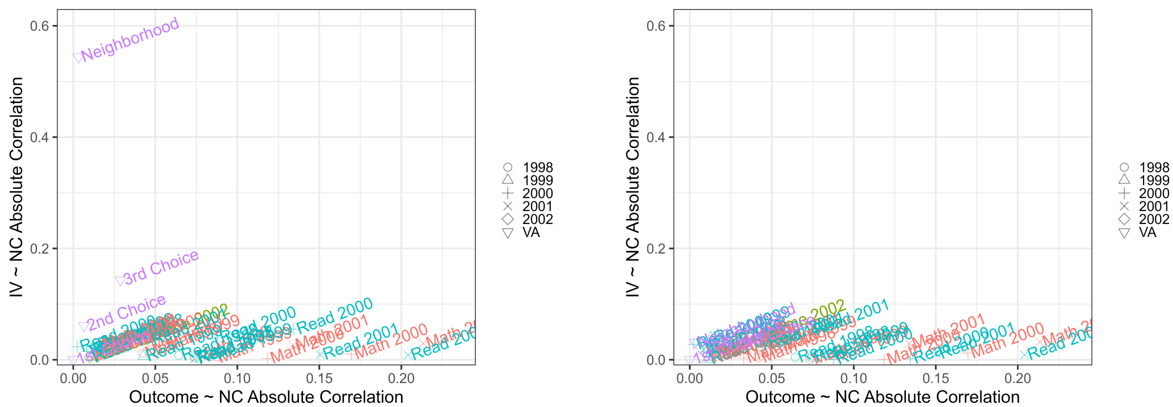
Figure 4: Correlations of Different Negative Control Variables with the Instrument and the Outcome in Autor et al. (2013)




The figure shows a scatter plot of the absolute value of the correlation of different negative controls with the instrument (on the x-axis) and the outcome (on the y-axis). The negative controls, the instrument, and the outcome were first residualized by regressing them on all control variables. Each observation is one negative control. Note that the axis scales are different. For presentation purposes, negative controls are grouped into categories denoted by color and shape. Labels are shown for specific negative controls that exhibit a high correlation with both the outcome and the instrument. Except for past outcomes, which were measured in 1970 and 1980 (unlabeled) as well as in 1990 (labeled "1990"), all other negative controls were measured in 1990. Source: authors' calculations using data from Autor et al. (2013).

Figure 5: Correlations of Different Negative Control Variables with the Instrument and the Outcome in Deming (2014)

(A) Original Instrument (B) Lottery Instrument




The figure shows a scatter plot of the absolute value of the correlation of different negative controls with the instrument (on the x-axis) and the outcome (on the y-axis). In Panel A, the instrument is the same instrument used by Deming (2014) (see text). In Panel B, the instrument is the binary lottery outcome for whether the student won the lottery. The negative controls, the instrument, and the outcome were first residualized by regressing them on all control variables and lottery fixed-effects. Each observation is one negative control. For presentation purposes, negative controls are grouped into categories denoted by color and shape. VA stands for school value-added; 1st, 2nd, and 3rd Choice are students' submitted preferences. Neighborhood is the student's neighborhood school. Source: authors' calculations using data from Deming (2014).

Table 1: Variable definitions and negative controls used in the analyses of Autor et al. (2013) and Deming (2014).

| | Autor et al. (2013) | Deming (2014) |
|---|---|---|
| Outcome (Y) | Changes in US commuting-zone manufacturing employment (stacked differences between 1990–2000 and 2000–2007) | Year 2003 test results |
| Endogenous Variable (X) | Imports from China (changes in U.S. import from China by industry between 2000-2007, weighted by commuting zone initial share of that industry) | Value-added (VA) of school attended |
| Instrument (IV) | Changes in Chinese imports by other high-income countries between 2000-2007, weighted by commuting zone initial share of that industry | 1. The instrument used in the paper; 2. The binary lottery result (Lottery) |
| Controls | Same as the original paper (see Table 2 ) | Same as the original paper (Lottery fixed effects, Year 2002 math and reading tests: values, squared, and missing) |
| Negative Controls (NC) | Predetermined economic characteristics of the commuting zone (see appendix C.1 for a list) | Years 1998-2001 math and reading results, year 2002 test results (past outcome), and all VA variables (choices 1-3, home VA). Total of 37 NCs. |

Table 2: Negative-Control Tests for the Instrument in Autor et al. (2013)

| p-values | Instrument: (Δ imports from China to other countries)/worker *Specification* (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| **Method:** | | | | | |
| F-test | 0.046 | <0.01 | <0.01 | <0.01 | <0.01 |
| Bonferroni | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 |
| Single Negative Control (1970 Outcome) | 0.246 | 0.828 | 0.683 | 0.747 | 0.709 |
| **Negative Control Variables:** | | | | | |
| | | | pre-1990 variables | | |
| **Control Variables:** | | | | | |
| % of Employment in Manufacturing | Yes | Yes | Yes | Yes | Yes |
| % College Educated | | | Yes | | Yes |
| % Foreign Born | | | Yes | | Yes |
| % Female Employment | | | Yes | | Yes |
| % Employment in Routine Occupations | | | | Yes | Yes |
| Average offshorability index of occupations | | | | Yes | Yes |
| Census Division Dummies | | Yes | Yes | Yes | Yes |

*Notes:* The table shows $p$-values from negative-control tests that attempt to predict the main instrument in Autor et al. (2013) using negative controls. Each row shows results for a different test approaches: F-test, Bonferroni-adjusted t-tests, and t-test from a single negative control (1970 outcome). Following the same combinations of control as the original study, each column shows results for a different set of controls. In all tests, we used the same set of negative controls, which include 52 lagged variables from the original study data that were measured before 1990 and not used as controls. The null assumption is that the instrument is uncorrelated with the negative controls. A $p$-value close to zero represents a rejection of this null. The sample includes 722 commute zones. Regressions were weighted by commute zone size, using the same weights as the original study. See Appendix C.1 for sample and variable definitions. Source: authors' calculation using replication data from Autor et al. (2013).

Table 3: Negative-Control Tests for the Instrument in Deming (2014)

| $p$-value | IV | |
|---|---|---|
| | Original | Lottery |
| | (1) | (2) |
| F-test | <0.01 | 0.648 |
| Bonferroni | <0.01 | 0.101 |

This table shows $p$-values from negative-control tests that attempt to predict the instrument in Deming (2014) using negative controls. Column (1) tries to predict the main instrument used in the paper. Column (2) uses a binary variable for the school lottery outcome. Each row shows results for a different testing approach: F-test and Bonferroni-adjusted t-tests. We Follow the same combinations of control as the original study. In all tests, we used the same set of negative controls, which include lagged test scores and school value added for the neighborhood school and the first three choices. The negative controls are taken from the original study data. The null assumption is that the instrument is uncorrelated with the negative controls. A $p$-value close to zero represents a rejection of this null. See Appendix C.1 for sample and variable definitions. Source: authors' calculation using replication data from Deming (2014).

# A Additional Theory and Proofs

## A.1 Negative Controls for Instrumental Variable Specifications that Include Control Variables

In this section, we prove Theorem 1 when controls $C$ are included in the analysis. First, we present the independence and exclusion restriction assumptions when controls are included.

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, $P[y(x)|u]$ is a shorthand for $\Pr[Y(x) = y(x)|U = u]$.

**Assumption 1a.** *Independence.* $Z \perp\!\!\!\perp Y(x,z)|C$ *for all possible* $x, z$ *values.*

**Assumption 2a.** *Exclusion restriction.* $P(Y(x,z) = Y(x,z') = Y(x)|C = c) = 1$ *for all possible* $x, z, z', c$ *values, and for all* $i = 1, ..., n$.

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

We also adapt the definitions of APV and negative controls as follows.

**Definition 1a** (Alternative-Path Variable)**.** *A random variable* $U$ *is an alternative-path variable conditionally on a set of controls* $C$ *if there exist a vector of variables* $V$ *such that*

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

**Definition 2a.** *NC is a negative control if there exists an APV such that if* $Z \perp\!\!\!\perp U|C$ *then* $Z \perp\!\!\!\perp NC|\ U, C$

We are now ready to state a version of Theorem 1 with controls and present its proof (which also covers the case no control are used by letting $C$ to be degenerate.)

**Theorem 2.** *If* $NC$ *is a variable that satisfies the negative-control assumption (see Definition 2a) and* $Z \not\perp\!\!\!\perp NC|C$ *then* $Z \not\perp\!\!\!\perp Y(x)|C$

*Proof.* $Z \not\perp\!\!\!\perp NC|C$ implies that $Z \not\perp\!\!\!\perp U|C$. Else, if $Z \perp\!\!\!\perp U|C$ then by the negative-control assumption (see Definition 2a) $Z \perp\!\!\!\perp NC|\ U, C$. Based on auxiliary Lemma 2 (in the next subsection), $Z \perp\!\!\!\perp NC|\ U, C$ and $Z \not\perp\!\!\!\perp NC|C$ imply that $Z \not\perp\!\!\!\perp U|C$, a contradiction.

From direct instrument link we get that $Z \not\perp\!\!\!\perp U|C$ implies $Z \not\perp\!\!\!\perp U|V, C$. Then by path indication, this implies that $Z \not\perp\!\!\!\perp Y(x)|V, C$. Finally, from V-validity we get that

$$Z \not\perp\!\!\!\perp Y(x)|C$$

□

## Auxiliary Lemmas

**Lemma 1.** *Let $A, B, Q, C$ be four (possibly vectors of) random variables. If $A \perp\!\!\!\perp B|Q, C$ and $B \perp\!\!\!\perp Q|C$ then $A \perp\!\!\!\perp B|C$.*

*Proof.* Because $A \perp\!\!\!\perp B|Q, C$ we have for all $a, b, q, c$

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

where the last line follows from the conditional independence of $B$ and $Q$, conditionally on $C$. Integrating over the distribution of $Q|C$ we get

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

and therefore

$$P(a, b|c) = P(a|c)P(b|c)$$

□

**Lemma 2.** *If $A \perp\!\!\!\perp B|Q, C$ and $A \not\perp\!\!\!\perp B|C$ then $A \not\perp\!\!\!\perp Q|C$ and $B \not\perp\!\!\!\perp Q|C$*

*Proof.* Assume by contradiction that $B \perp\!\!\!\perp Q|C$. Therefore, by Lemma 1, because $A \perp\!\!\!\perp B|Q, C$ we have that $A \perp\!\!\!\perp B|C$ which contradicts the assumption. Similarly for $A \perp\!\!\!\perp Q|C$. □

**Lemma 3.** *If $A \perp\!\!\!\perp B$, $A \perp\!\!\!\perp C$ and $A \perp\!\!\!\perp B|C$ then $A \perp\!\!\!\perp C|B$*

*Proof.* For every $a, b, c$ values of $A, B, C$ $P(a|b) = P(a) = P(a|c) = P(a|b, c)$ □

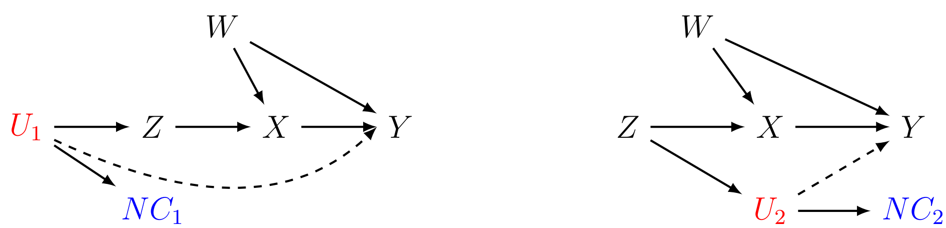
**Lemma 4.** *If $A \perp\!\!\!\perp B|C$ and $A \not\perp\!\!\!\perp C|B$ then either $A \not\perp\!\!\!\perp B$ or $A \not\perp\!\!\!\perp C$*

*Proof.* Assume by contradiction that both $A \perp\!\!\!\perp B$ and $A \perp\!\!\!\perp C$. Then by Lemma 3 $A \perp\!\!\!\perp C|B$ which contradicts the premise. □

## A.2 Negative Control Exposures

In biostatistics and epidemiology, negative control exposures (NCEs) are variables known not to affect the outcome but have the same suspected confounding mechanism as the treatment-outcome relationship (Shi et al., 2020). NCEs are useful for cases where the researcher knows the treatment is associated with some suspected confounder, however, it is not clear if the outcome is also associated with it. Assuming the NCE suffers from the same potential confounding mechanism (and only this source of confounding), researchers can test if the NCE is independent of the outcome, conditional on the treatment. The independence test should be conditional, as the NCE and outcome are already linked through the treatment. Rejection of this conditional independence implies that the treatment and the outcome are also related through a confounder.

NCEs can also apply to the context of instrumental variables. NCEs are useful when the instrument is associated with another variable, yet it is not clear if the other variable is related to the outcome. Figure A1 demonstrates this using two DAGs. In both panels, the outcome $Y$ is associated with the NCE, regardless of the instrument validity. However, the outcome will also be linked with the NCE conditional on the instrument $Z$ only if the instrument is invalid.




(A) $NC_1 \not\perp\!\!\!\perp Y|Z$ implies $U_1 \longrightarrow Z$, violating instrument independence.

(B) $NC_2 \not\perp\!\!\!\perp Y|Z$ implies $Z \longrightarrow U_2$, violating instrument exclusion.

Figure A1: Negative Controls Exposures for Instrumental Variables: Graphical Illustrations

*Notes:* The figure illustrates the use of negative control exposure variables for evaluating the validity of instrumental-variable designs. In both panels, $X$ is the endogenous variable, $Y$ is the outcome, $Z$ is the instrument, and $W$ is a potential confounder that motivates the use of the instrument. The variables $U_1$ and $U_2$ are unobserved *APV-exposure* that pose threats to identification, in the sense that if they are related to the outcome $Y$ (through the links specified by a dashed arrow), either independence (Panel A) or the exclusion restriction (Panel B) is violated. In each case, an observed *negative control exposure* ($NC_1$ or $NC_2$) that is related to the respective APV-exposure ($U_1$ or $U_2$) but not directly to $Y$ can be used to evaluate the presence of the problematic link.

# B Examples and Counterexamples

## B.1 Non-Causal APV

In the example presented in Figure A2, $U_2$ is a valid APV, which satisfies path indication, even though there is no causal link from it to $Y$. Path indication is satisfied because both $U_2$ and $Y$ are commonly caused by $U_1$. For instance, consider $Z$ to be teacher assignment that is claimed to be quasi-random, $X$ the teacher value added of the teacher and $Y$ some test scores. The variable $W$ represents the usual concern that some students are able to switch classrooms and end up with different teachers. In this example, $U_1$ is unobserved ability, which directly affects test scores. $U_2$ represents detailed test scores in some previous exams that are also unobserved. The dashed arrow represents a concern that principals allocate students to teachers based on the detailed test score (e.g., students with low math scores are assigned to a specific teacher).

In this example, $U_2$ satisfies path indication even though it is not causally affecting the outcome as detailed past test scores do not directly affect future test scores. However, there is a path between the previous detailed test scores and the current test scores, because both are affected by ability ($U_1$).

The variable $NC$ is aggregated previous test scores, which averages past scores in math with other subjects. In this setting, $NC$ is a negative control, with $U_2$ as an APV. A correlation between the instrument and aggregated lagged test scores would imply an alternative path from the instrument to the outcome. Specifically, this will violate independence as students with different abilities would sort into different teachers based on their previous math scores.

Of note is that in this scenario, $U_1$ is also an APV. However, $NC$ is a valid negative control with respect to $U_2$ but not with respect to $U_1$ alone, as conditionally on the unobserved ability, there is still a correlation between the negative control and the instrument ($Z \not\perp NC | U_1$), i.e, the teacher assignment is not independent of the aggregated test scores.

## B.2 Violation of Path Indication: Multidimensional Variable

We now discuss a case where $U$ is a vector of dummy variables. For instance, assume $Z$ is the teacher assignment, which is claimed to be quasi-random, $X$ is the teacher value added of the teacher and $Y$ is the test scores. The variable $W$ represents the usual concern that some students are able to switch classrooms and end up with different teachers. Assume that $U$ is the student's hobbies, which they report to the school, but unobserved to the researcher. Each student can have more than one hobby. We plot this example in the DAG in Figure

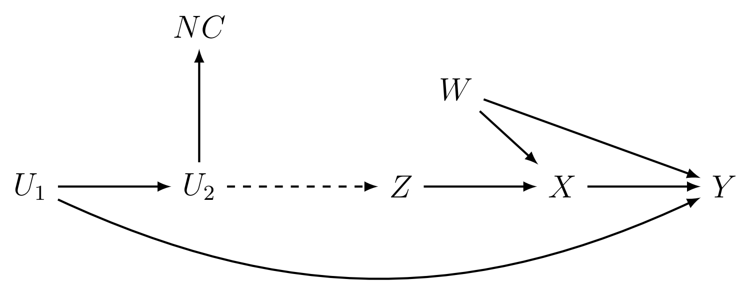



Figure A2: An illustration of causal and non-causal APVs.

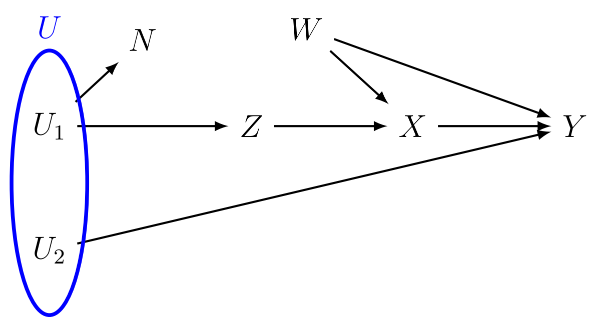



Figure A3: Violation of APV conditions when $U$ has multiple components. The ellipse around $U_1, U_2$ is a visual marking of taking $U = (U_1, U_2)$.

A3.

Assume that $U_1$ is having basketball as one of your hobbies. Assume also that basketball is correlated with the instrument. For example, one teacher also teaches basketball and so the basketball players (who list basketball as their favorite hobby) are automatically assigned to her. However, basketball does not affect test scores. Let $U_2$ indicate having math as one of your hobbies. Assume that students that report math as their favorite hobby tend to perform better in exams and that math lovers are randomly allocated across teachers. Finally, assume that $U_1 \perp\!\!\!\perp U_2$.

In this case, $U$ is not an APV. The vector $U$ does satisfy latent instrument validity. However, it does not satisfy path indication. The instrument is valid ($Z \perp\!\!\!\perp Y(x)$) as even though the instrument is correlated with having basketball as a hobby, this is irrelevant for test scores.

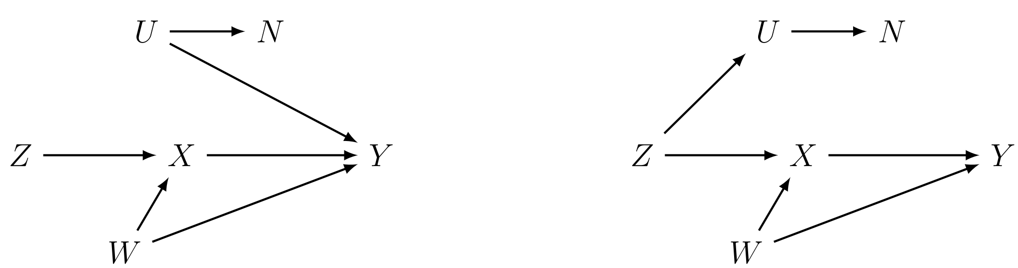
Defining $N$ as participating in an after-school basketball program, we find that $N$ is not a proper negative control. Even though $Z \perp\!\!\!\perp N|U$, it is still not a negative control since $U$ is not an APV.

## B.3 Heterogeneity-Based Violation of Path Indication

Figure A4 describes a random variable $U$ that is linked to both the instrument $Z$ and the potential outcome $Y(x)$, yet is not an APV since it does not satisfy path indication. Hence the variable marked as $N$ in Figure A4, is not a valid negative control variable. In Figure A4, the instrument affects $U$ for males but not for females. On the other hand, $U$ affects $Y$ for females but not for males. Path indication is not satisfied since even though the instrument is valid ($Z \perp\!\!\!\perp Y(x)$), it is still linked with $U$ ($Z \not\perp\!\!\!\perp U$).

In this example, for a large enough sample size we will conclude that $Z \not\perp\!\!\!\perp N$ (due to the affect among males), but from this test, one cannot deduce that the instrument is invalid since $U$ is not an APV.

A remark is that had we considered an analysis among males only, path indication would still be violated, because $U$ is not associated with $Y(x)$. However, among females only, $U$ is an APV, $N$ is a valid negative control, and the instrument would pass the test (among the females, $Z \perp\!\!\!\perp N$).




(A) Females (B) Males

Figure A4: Illustration of a violation of V-validity via DAGs

## B.4 Violation of $V$-validity: Common Effect of the Outcome and the Suspected APV

Figure A5 presents a situation with no valid APV. We examine $U$ as a candidate APV, and consider $V$ in the DAG as the potential $V$ in Definition 2. We see that latent instrument validity holds: while $Z \not\perp\!\!\!\perp Y(x)|V$, because $V$ is a common effect (a collider) of $U$ and $Y$, controlling for $U$ in addition to $V$ blocks the flow of association (Pearl, 2009) resulting in $Z \perp\!\!\!\perp Y(x)|U, V$. Path indication holds trivially because $Z \not\perp\!\!\!\perp Y(x)|V$. Direct instrument link also holds because of the effect of $U$ on $Z$. However, it is clear $U$ should not be an APV. An association between $Z$ and $U$ does not imply the instrument is invalid. Here comes $V-$validity to the rescue. In Figure A5, the instrument is valid ($Z \perp\!\!\!\perp Y(x)$), 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 choice for $V$ exists to satisfy Definition 2. Therefore $U$ is not an APV, and the random variable $N$ is not a negative control.

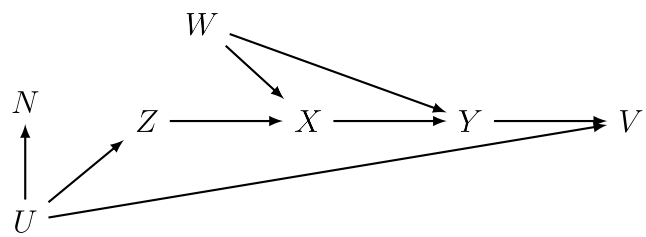



Figure A5: Violation of APV conditions when $U$ the instrument is valid, and $U$ is not an APV because there is no $V$ satisfying all four conditions.

### B.5 A Vector of Negative Controls Which is Not a Negative Control

Let $R_1, R_2$ be two i.i.d. random variables who are distributed Bernoulli with $p_1 = p_2 = 0.5$. Let $U$ be some random variable, independent of $R_1, R_2$. Let $Z$ be the instrumental variable, and assume that it can be written as $Z = \alpha(R_1 \oplus R_2) + \beta U + \epsilon_z$. Assume that $Y(x) = x + U + \epsilon_y$, such that $U$ is an APV. The instrument is valid if $\beta = 0$.

Assume 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 $Z \perp\!\!\!\perp NC_i|U$. 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 $Z \not\perp\!\!\!\perp (NC_1, NC_2)|U$ whenever $\alpha \neq 0$. Therefore $(NC_1, NC_2)$ does not satisfy the negative control assumption and so it is not a negative control. Indeed, even if the instrument 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 instance, 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, $Z \not\perp\!\!\!\perp NC_2$ and $NC_2$ will not satisfy the negative control assumption.

## C Details of the Implementation of Negative-Control Tests Using Data from Prior Studies

This section provides additional details about our application of our proposed method to the instrumental variable design used in Autor et al. (2013) and Deming (2014) (discussed in Section 4). We are grateful to the authors of these prior works for making these analyses possible by publicly posting their data and code. In each case, we first used the publicly posted data to replicate the related original study results (this step is not further discussed here). We then applied our method, which attempts to predict the instrument using negative control variables from the original study data.

## C.1 Example I: Shift-Share Instruments

**Sample construction.** For this analysis, we use the original study data from Autor et al. (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 replication data, including the original instrument, controls, and additional variables not used in the original study, some of which we use as negative controls in our current analysis. Our analysis uses the same control variables that were used in the original analysis. The controls we used are from the year 2000. The outcome, exposure, and instrument are differences between the years 2007 and 2000. Controls are the lagged year 2000 values. We use all other variables in the ADH data set that were measured in or before 1990 (excluding the instrument and controls) as negative controls. Note that ADH also used another version of the instrument, measured between 1990–2000. We do not evaluate this version, because we do not observe variables that were measured early enough before 1990 to serve as negative controls.

**Original Falsification Tests** ADH conducted falsification exercises to evaluate the concern that a secular decline in US manufacturing might be the underlying reason for the increased imports from China. To that end, 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, leading the authors to conclude that this is unlikely to be a major concern. We replicated this analysis and obtained a similar result. This original exercise is similar in spirit to our proposed approach, although it is using these different negative controls separately and not jointly, as we propose. The rest of this section discusses additional falsification tests which we performed using alternative negative control variables.

**Methods.** The main step in our analysis is testing for conditional independence of the original instrument and different negative controls, with adjustment for different sets of control variables that were used in the original study.

We repeat this exercise for each of the five different sets of control variables that were used in the original study and detailed in columns 2–6 of Table 3 of ADH. We weighted all regressions using the same weights used by ADH in the original study (`timepwt48`).

## School Choice Lotteries

**Sample construction.** Our second replication analyzes data from Deming (2014). We use the study data, originally sourced from a public school choice lottery in Charlotte-Mecklenburg (CMS). In these data, we observe the original study outcome, instrumental and control variable.

**Main variables** VAM calculation was done from the 1996–1997 through 2001–2002 school years. The outcome in the second stage is the school year 2002-2003. In the code, each school year is called by the year in which it ends, for example, school year 2002-2003 is called 2003.

In the article, the endogenous treatment is the value added (VAM) of the student's school. Based on the replication code, the instrument can be written as

$$IV_i = W_i VAM_i^1 + (1 - W_i)\, VAM_i^N \tag{8}$$

where we denote the binary school lottery outcome by $W$, the value added of the first-choice school by $VAM^1$ and the value added of the default neighborhood school by $VAM^N$. These original variables are included in the study replication data.

We use Deming's variants of the school value-added estimates that are posted in the replication data. Specifically, the original paper used different variants of the VAM estimates that differ across three dimensions: 1) the VAM model specification (equation 1 in the original study): mixed-effects, fixed-effects, or AR; we use the posted version, which only includes mixed-effects specifications. 2) the controls used for VAM calculations; we use the posted version, which includes past grades.[16] 3) the number of lagged years of the grades used; we use the maximum number, which is four. We also construct two alternative instruments. First, the raw lottery outcome:

$$Z_i^1 = W_i.$$

[16] The richer specifications (models 3-4 in the original study) include personal information that was not made publicly available.

Second, a product of the lottery and the first-choice school,

$$Z_i^2 = W_i VAM_i^1.$$

As negative controls, $NC_{ijt}$, we use lagged test scores from the school year 2000-2001, which is two years before the lottery. We also used the VAM of the 3 schools that the student applied to in the lottery, and the neighborhood school's VAM.[17]

[17] Note that only lagged test scores from year $-1$ are used as controls (in both the original study and in our replication), so there is no overlap between these controls and the earlier test score we use as negative controls.