Characterizing the Instrumental Variable Identifying Assumption as Sample Selection Conditions

Christian Belzil
Jörgen Hansen

DISCUSSION PAPER SERIES

Characterizing the Instrumental Variable Identifying Assumption as Sample Selection Conditions

Christian Belzil
*Ecole Polytechnique,
ENSAE and IZA*

Jörgen Hansen
*Concordia University,
CIRANO, CIREQ and IZA*

ABSTRACT

Characterizing the Instrumental Variable Identifying Assumption as Sample Selection Conditions[*]

We build on Rosenzweig and Wolpin (2000) and Keane (2010) and show that in order to fulfill the Instrumental variable (IV) identifying moment condition, a policy must be designed so that compliers and non-compliers either have the same average error term, or have an error term ratio equal to their relative share of the population. The former condition (labeled *Choice Orthogonality*) is essentially a no-selection condition. The latter one, referred to as *Weighted Opposite Choices*, may be viewed as a distributional (functional form) assumption necessary to match the degree of selectivity between compliers and noncompliers to their relative population proportions. Those conditions form a core of implicit IV assumptions that are present in any empirical applications. They allow the econometrician to gain substantial insight about the validity of a specific instrument, and they illustrate the link between identification and the statistical strength of an instrument. Finally, our characterization may also help designing a policy generating a valid instrument.

JEL Classification: B4, C1, C3

Keywords: instrumental variable methods, implicit assumptions, treatment effects

Corresponding author:

Christian Belzil
Ecole Polytechnique
Département d'Economie
91128 Palaiseau Cedex
France
E-mail: christian.belzil@polytechnique.edu

---

1 Introductory Remarks

The analysis presented in this paper discloses a relationship between IV estimation and sample selection models that goes deeper than the role of the first-stage model within the comparative advantages (LATE) approach (Vytlacil, 2002, Imbens and Angrist, 1994, Heckman and Vytlacil , 2005). We build on ideas exposed in Rosenzweig and Wolpin (2000) and Keane (2010) and show that the IV identifying moment condition, usually expressed as a treatment-control orthogonality condition, may be re-interpreted as two mutually exclusive sample selection conditions that jointly restrict (i) differences in error terms between those affected (compliers) and those unaffected (non-compliers), (ii) and their relative shares (the noncomplying odds ratio).

To fulfill the moment condition, a policy must be designed so that compliers and non-compliers either have the same average error term, or have an error term ratio equal to their relative share of the population. The former condition, which we label *Choice Orthogonality*, is essentially a no-selection condition. The latter one, which we refer to as *Weighted Opposite Choices,* may be viewed as a distributional (functional form) assumption necessary to match the degree of selectivity between compliers and non-compliers to their relative population proportions.

Our characterization is valid for virtually all micro-econometric applications of IV found in the empirical literature, as long as a strictly positive fraction of individuals remain unaffected by a policy, and as long as the parameter of interest is identified by a moment (orthogonality) condition.[1] It is valid in the classical case where the error term of the outcome equation is additive as well as in a comparative advantages framework, or in the LATE framework. Unlike Rosenzweig and Wolpin (2000) and Keane (2010), our approach does not require to rely on a tightly specified data generating process. All is needed is an outcome equation. Nevertheless, those conditions form a core of implicit IV assumptions that are present in any empirical applications. They are valid in the classical linear IV framework, as well as in any multiplicative error term framework (such as the correlated random coefficient model).

Our result is constructive. Transforming the treatment-control orthogo-

---

[1]Although this condition may sound innocuous, anecdotal evidence suggests that many empiricists using IV methods disregards identification and do not distinguish between estimators and estimands.

nality condition into sample selection conditions, allows the econometrician to gain substantial insight about the validity of a specific instrument, and may also help design a policy intervention that generates a valid instrument. Essentially, any econometrician who has some priors about the identity of those affected by a policy and their relative shares of the total population, can use our characterization of the implicit IV assumptions to evaluate the likelihood that the identification condition be fulfilled. Finally, our result sheds light on the relationship between the statistical strength of an instrument (the weak-strong instrument paradigm), and its capacity to achieve identification.

In the next section, we briefly review some recent literatures on IV estimation so to clarify the specificities of our analysis. Section 3 contains the main core of the paper, and is devoted to the re-expression of the IV identifying assumption. In the following section, we introduce the distinction between Dictatorial and Incentive-Based Policy Interventions. In Section 5, we discuss a few related practical issues and discuss the extension of the outcome equation to more general forms. The conclusion is found in Section 6.

2   Background Material

In the recent econometric literature, two distinct types of implicit IV assumptions have been analyzed. First, Heckman (1997), Vytlacil (2002), Heckman and Vytlacil (2005), and Heckman, Urzua and Vytlacil (2007) have analyzed potential misspecifications of the first stage model of an IV regression within a correlated random coefficient model framework (Imbens and Angrist, 1994). As is well illustrated in Heckman, Urzua and Vytlacil (2007), and unlike for a classical common slope IV regression, the first stage regression model chosen by the econometrician hides some crucial implicit behavioral assumptions. When those assumptions are unfounded, the "Monotonicity" property may be violated, and IV may produce uninterpretable results.[2] These findings have also generated a substantial number of papers that debate the rele-

---

[2]This is naturally explained by the fact that different instruments may generate different population treatment effects. For instance, mis-specification may culminate in parameter estimates of opposite signs of their population counterparts (see Heckman and Vytlacil, 2005).

4

vance of estimating Local Average Treatment effects.[3]

In parallel, many economists have pointed out the behavioral implications underlying the IV identifying orthogonality condition. This constitutes the second type of implicit assumptions that have been considered in the literature, and is the focus of our paper.[4]

In an important contribution, Rosenzweig and Wolpin (2000) have examined a wide sets of microeconomic models (in Labor and Development), and provided a "case by case" analysis of the implicit assumptions required for interpreting IV estimates reported in the existing literature. In order to develop their argumentation, Rosenzweig and Wolpin analyze specific examples in which the econometrician's error term is decomposed into model specific components, and show how implementing policy changes is generally not sufficient to generate orthogonality conditions. This line of reasoning has also been followed by Keane (2010).[5]

To be clear, the objective of our paper is to build on Rosenzweig and Wolpin (2000) and Keane (2010) and show how their results can also be cast within standard sample selection arguments, without relying on specific data generating processes. For this reason, our approach also displays similarities with the literature on IV estimation of the correlated random coefficient regression model and the LATE approach. However, it should be clear that our analysis is for any IV identifying conditions and is therefore not confined to models where outcome equations are subject to comparative advantages or heterogenous treatment effects.

---

[3]See Heckman, Urzua and Vytlacil (2005) and Imbens (2009) for opposite views about the relevance of the LATE parameter.

[4]The identifying moment condition that we analyze in this paper, is actually called the "Independence" assumption in Imbens and Angrist (1994) and Heckman and Vytlacil (2005).

[5]Keane (2010) uses the term "atheoretical" approach to economics to refer to the empirical literature using IV methods. Belzil (2007) uses similar arguments in his survey of the structural literature on returns to schooling.

3 Re-interpreting the IV Identifying Condition

Our main contribution is to produce a general characterization of the IV identifying assumption that is portable across any empirical application. We show that, as long as a strictly positive fraction of the population is not affected by an intervention, substantial insight may be gained by introducing the complying status into the identifying condition. There are at least four good reasons to do it.

First, most realistic policy interventions set in a context where individuals choose among a discrete set of options, can only affect a subset of the population.

Second, most policies used in empirical work imply a non-trivial degree of selectivity between compliers and non-compliers.[6]

Thirdly, in the applied treatment effect literature, the most favored interpretation of IV estimates (the LATE parameter) is itself explicitly based on the complying status.[7] So, there is no logical reason why the identifying moment condition would not take into account distinctions between those affected and those who are not.

Finally, the fourth reason is that economists implementing an IV approach are bound to have priors about the identity of those affected and unaffected, and possibly about their relative shares of the total population. They are also bound to have some idea (precise or imprecise) about what unobserved choices could contaminate the identifying assumption.

3.1 The Outcome Equation

We consider an econometrician who wants to estimate the effect of a right-hand side endogenous variable, denoted S, on a left-hand side variable Y.

Assumption 1- The Outcome Equation

We assume that the true model is given by the following expression:

---

[6]This is the case, for instance, in the schooling literature where policies such as compulsory schooling regulation changes are often used as instrument.

[7]For instance, the LATE interpretation of IV requires a positive proportion of individuals who are not affected by a given policy change.

$$Y_i = \alpha + \beta \cdot S_i + \delta \cdot W_i \tag{1}$$

where both S and W are endogenous variables, β is the parameter of interest, and α is the intercept term.

We use the common slope version of the wage equation so to eliminate ambiguity regarding which treatment effect should (or could) be targeted. All the analysis presented below would carry through even if individuals were endowed with their own β. All is required is that β, be identified by a moment condition.

Assumption 2- Model Misspecification

To motivate our approach, we assume that the econometrician does not observe W_i, and must therefore work with a mis-specified (or incompletely specified) model.[8] When estimating the model, the econometrician assumes that

$$Y_i = \alpha + \beta \cdot S_i + \varepsilon_i \tag{2}$$

where ε_i is actually equal to $\delta \cdot W_i$ and is the econometrician's error term.

At this stage, we do not give a precise set up of the model that generates S and W. For instance, S and W may be chosen simultaneously, or sequentially. All we assume is that both S and W are endogenous choice variables. Throughout the paper, we refer to them as consumption levels, and assume that they are driven by a common unobserved (to the econometrician) component, θ_i, with distribution function $H(.)$. In this framework, we think of θ_i as the individual specific model primitives (tastes, preferences,.etc.).

This framework fits a large number of empirical applications considered in Labor Economics, Health Economics, Public Economics, Crime Economics, or other areas of applied microeconomics. For instance, in the human capital literature, one may be interested in estimating the effect of education (S) on earnings (Y) in a context where post-schooling skills (W) are unobserved. In Health economics, the effect of pollution exposure (S) on birth weight (Y) may be investigated in a context where the econometrician does not observe alcohol (or drug) consumption of the pregnant woman (W).

---

[8]This is easily motivated by the common belief that all empirical models are (to some extent) misspecified, and likely to omit the effect of unobserved individual choices.

3.2 The Instrument

As a starting point, assume that the econometrician wants to estimate β by IV methods. The instrument, Z, is a binary indicator such that $Z_i = 1$ for treatment, $Z_i = 0$ for control, and which records an unanticipated policy change. Without loss of generality, we assume that the econometrician has access to a single instrument.

Assumption 3- Ex-Ante Randomization

As do all empirical economists using IV methods, we only consider policy interventions that are subject to *Ex-Ante Randomization*. That is the distribution of individual fixed endowments (the model primitives) is invariant to policy exposure. Formally, we have that

$$H(\theta_i \mid Z_i = 1) = H(\theta_i \mid Z = 0) \tag{3}$$

Using standard notation, the IV estimator is defined as

$$\beta_{IV} = (\sum_i Z_i S_i)^{-1} \cdot \sum Z_i Y_i \tag{4}$$

To be consistent with the empirical literature, we assume away any general equilibrium effects.

We now introduce a random variable, C_i, which records the complying status. Precisely, we have

$$
\begin{aligned}
C_i &= \quad 1 \text{ if } S_i(Z_i = 1) \neq S_i(Z_i = 0) \text{ (Compliers)} \\
C_i &= \quad 0 \text{ if } S_i(Z_i = 1) = S_i(Z_i = 0) \text{ (Non-compliers)}
\end{aligned}
$$

Assumption 4- Strictly Positive Non-Complying Probability

We only consider policy interventions that leave a strictly positive fraction of the treatment population unaffected. That is, we assume that $\Pr(Z_i = 1, C_i = 0) > 0$.

To simplify notation, we denote $\Pr(Z_i = 1, C_i = 0)$ by $\Psi_0(Z_i = 1)$, and $\Pr(Z_i = 1, C_i = 1)$ by $\Psi_1(Z_i = 1)$, and denote the ratio of those probabilities by $\Psi_{0,1}(Z_i = 1)$. That is

$$\frac{\Pr(Z_i = 1, C_i = 0)}{\Pr(Z_i = 1, C_i = 1)} \equiv \frac{\Psi_0(Z_i = 1)}{\Psi_1(Z_i = 1)} \equiv \Psi_{0,1}(Z_i = 1) \tag{5}$$

The term, $\Psi_{0,1}(Z_i = 1)$, which we call the *Noncomplying Odds Ratio*, is the large sample equivalent of the observed relative shares of compliers and non-compliers in the data.[9]

3.3 Characterizing the Implicit Assumptions

Because the treatment-control indicator is binary (0,1), identification of the parameter of interest, β, requires that for each observation i,

$$E(Z_i \varepsilon_i) = \Pr(Z_i = 1) \cdot E(\varepsilon_i \mid Z_i = 1) = 0 \qquad (6)$$

Because the treatment group is composed of both compliers (for whom $C_i = 1$), and noncompliers (for whom $C_i = 0$), (5) may be decomposed further. That is

$$
\begin{aligned}
E(Z_i \varepsilon_i) &= \Pr(Z_i = 1) \cdot E(\varepsilon_i \mid Z_i = 1) \\
&= \Psi_1(Z_i = 1) \cdot E(\varepsilon_i \mid Z_i = 1, C_i = 1) + \Psi_0(Z_i = 1) \cdot E(\varepsilon_i \mid Z_i = 1, C_i = 0) \\
&= 0
\end{aligned}
\qquad (7)
$$

We now show that from a purely economic standpoint, it is informative to decompose (7) into two mutually exclusive conditions.

3.3.1 Case 1: *Choice Orthogonality (with Respect to Complying Status)*

Choice Orthogonality arises when those affected and those unaffected share (on average) the same ε. This is the case when

$$E(\varepsilon_i \mid Z_i = 1, C_i = 1) = E(\varepsilon_i \mid Z_i = 1, C_i = 0) = 0 \qquad (8)$$

Note that when *Choice Orthogonality* is fulfilled, neither $\Pr(Z_i = 1, C_i = 1)$ nor $\Pr(Z_i = 1, C_i = 0)$ play any role in the moment condition. So, as long as the condition is fulfilled, the *Noncomplying Odds Ratio* is irrelevant. In

---

[9]In a more general model in which agents are subject to stochastic shocks, the complying status is a probabilistic event from the perspective of the agent (conditional on type). Within our framework, it is sufficient for the complying status to be stochastic from the perspective of the econometrician.

words, *Choice Orthogonality* implies that a policy, which initially randomizes a population into treatment and control and affects a subset of the treatment population only, must also reproduce another form of randomization by splitting the population into two groups (compliers and non-compliers) that have the same average post-intervention choices (error term).

Now, consider the existence of a policy intervention which separates the population into two groups that have an identical heterogeneity distribution. If so, we say that *Post-Treatment Identity Randomization* is fulfilled. That is the case when

$$H_1(.) = H_0(.) \tag{9}$$

As should be clear, *Post-Treatment Identity Randomization* is a notion that characterizes the set of individuals affected by a policy through policy invariant components of the model (individual heterogeneity) as opposed to choices.[10] In our simple framework in which W (or ε) depends on θ only (and not on C_i) , *Post-Treatment Identity Randomization* is a sufficient condition to generate *Choice Orthogonality*. This is not the case in more general models.[11]

3.3.2 Case 2: *Weighted Opposite Choices (with Respect to Complying Status)*

Obviously, *Choice Orthogonality* is not a necessary condition for the realization of the identifying condition. Differences between those affected and unaffected may be consistent with the fulfillment of the orthogonality condition. This leads us to the second case.

A policy is characterized by *Weighted Opposite Choices* when

$$E(\varepsilon_i \mid Z_i = 1, C_i = 1) \neq 0$$
$$E(\varepsilon_i \mid Z_i = 1, C_i = 0) \neq 0$$

and when

---

[10]In practice, very few policy interventions are likely to generate *Post-Treatment Identity Randomization.*

[11]It should also be clear that neither *Post-Treatment Identity Randomization,* nor *Choice Orthogonality* are implied by *Ex-Ante Randomization.*

$$\frac{E(\varepsilon_i \mid Z_i = 1, C_i = 1)}{E(\varepsilon_i \mid Z_i = 1, C_i = 0)} = -\Psi_{0,1}(Z_i = 1) \tag{10}$$

In words, (10) implies that behavioral differences between compliers and non-compliers are conditioned by the *Noncomplying Odds Ratio* $\Psi_{0,1}(Z_i = 1)$. From now on, we refer to $\frac{E(\varepsilon_i \mid Z_i=1, C_i=1)}{E(\varepsilon_i \mid Z_i=1, C_i=0)}$ as the *Error Term Ratio. Weighted Opposite Choices* implies that no matter what policy is considered, the *Error Term Ratio* is exactly equal to the relative shares of non-compliers and compliers in the treatment group.

As an example, a policy that affects only 10% of the population, leading to a *Noncomplying Odds Ratio* equal to 9, fulfills this condition only if those affected consume a quantity of good W which is 9 times the consumption of those unaffected.[12] A policy affecting 2% would require a consumption ratio equal to 49.

Policies that affect a more balanced subset of the population (those for which the odds ratio approaches 1) do not require such extreme differences in the *Error Term Ratio*, but are still restricted. For instance, a policy affecting 50% of the population fulfills *Weighted Opposite Choices* if average consumption of those affected is exactly equal to the opposite of the average consumption of those unaffected ($E(\varepsilon_i \mid Z_i = 1, C_i = 1 = -E(\varepsilon_i \mid Z_i = 1, C_i = 0)$).

There is however an important asymmetry between the *Error Term Ratio* and the *Noncomplying Odds Ratio*. While the former is informative about individual choices, the latter is an irrelevant quantity from the perspective of the agent. That is, in absence of general equilibrium effect, agents' choices would not react to any counterfactual manipulation of the *Noncomplying Odds Ratio*, unless their own θ has changed.[13] For instance, an agent i (endowed with θ_i), would exercise the same choice even if the cross-section distribution of θ is manipulated externally.

---

[12]In empirical work, and to our knowledge, empiricists never argue in favor of orthogonality conditions along those lines.

[13]It is also important to note that either *Choice Orthogonality* or *Weighted Opposite Choices* may be fulfilled even if individuals are not randomized ex-ante (with respect to individual specific heterogeneity). However, such policy interventions are practically never considered in the applied literature.

3.4 The IV Identifying Condition as Sample Selection Conditions

The analysis presented in this section discloses an interesting paradox. The current popularity of the IV approach is largely due to the willingness to depart from the parametric selection models of the 1980's.[14] Yet, the IV identifying condition relies on the prime objects of the literature on selection models, and the analogy between either *Choice Orthogonality* or *Weighted Opposite Choices* and classical sample selection models is striking.

Essentially, *Choice Orthogonality* is realized when there is no selection with respect to complying status. Indeed, both $E(\varepsilon_i \mid Z_i = 1, C_i = 1)$ and $E(\varepsilon_i \mid Z_i = 1, C_i = 0)$ are incidentally truncated means which play a central part in the selection literature (Heckman, 1979).

When it is impossible to maintain *Choice Orthogonality*, the econometrician must assume *Weighted Opposite Choices.* By doing so, the econometrician restricts the shape of the distribution of ε so to obtain a ratio of truncated mean error terms equal to the *Noncomplying Odds Ratio.* Put differently, the degree of selectivity tolerated by the IV moment condition imposes strong assumptions on the distribution of the error term.[15] So, in either case, the IV identifying condition boils down to a selection condition.

It is important to put our finding in perspectives with those papers that have already noted the key role that selection models play in interpreting IV estimates. In a seminal piece, Vytlacil (2002) has established that the LATE approach to IV estimation (Imbens and Angrist, 1994) is equivalent to a non-parametric selection model.[16] Our finding is different. It shows that the link between IV estimation and selection models goes much beyond the

---

[14]The author interested in the evolution of the microeconometric literature over the last 40 years should consult Heckman (2010).

[15]The issue is analogous to the characterization of the truncated means of a scalar random variable X with mean μ. To see this, note that

$$\begin{aligned} \mu &= E(X \mid X < c) \cdot \Pr(X < c) + \\ E(X &\mid X \geq c) \cdot \Pr(X \geq c) \end{aligned}$$

and that any restriction on μ automatically implies restrictions on the truncated means and the masses above and below a given truncation point c.

[16]To establish this result, Vytlacil needed to assumed independence (as do Imbens and Angrist, 1994, and Heckman and Vytlacil, 2005).

sole correlated random coefficient regression model. As noted earlier, the interpretation of the moment condition as sample selection conditions arise in the most standard framework, in which the error term of the outcome equation is additive, although it is obviously valid in a comparative advantages framework.[17]

4 The Distinction between Dictatorial and Incentive-Based Policy Interventions

Until now, the notion of policy intervention has remained vague. We now go a step further and link our theoretical results, to the qualitative nature of a policy intervention. It is informative to separate policy interventions into two distinct groups; namely *Incentive-Based Policy Interventions* and *Dictatorial Interventions*.

4.1 Incentive-Based Policy Interventions

We define *Incentive-Based Policy Interventions* as those that changes the incentive to consume a specific good, by modifying some parameters. For instance, a policy may offer a monetary incentive (or a disincentive) conditional on consuming a particular quantity of good S. In the schooling literature, any policy increasing the net utility of attending school (tuition subsidies, transportation subsidies, building new universities) would fall into this category.

4.2 Dictatorial Interventions

A *Dictatorial Intervention* affects individual decisions by restricting the choice set. Formally, it imposes a minimum consumption level, $s_{\min}$, or a maximum, $s_{\max}$. Compulsory schooling regulations often used in the schooling literature belongs to the class of dictatorial interventions.

---

[17]Finally, it should be reminded that those conditions have been derived under the assumption that there is a non-empty set of non-compliers. Obviously, in a case where the entire population is affected (that is when $\Psi_0(Z_i = 1) = 0$), we recover the usual treatment-control orthogonality condition, namely $E(\varepsilon_i \mid Z_i = 1) = 0$.

The distinction between *Incentive-Based Policy Interventions* and *Dictatorial Interventions* is not innocuous. With dictatorial experiments, it is possible to write the identifying condition as an expression that contains explicitly the mandatory consumption level itself. It is also interesting to note that the *Noncomplying Odds Ratio* of a dictatorial policy may be inferred directly from pre-intervention consumption.

For a minimum consumption policy, *Choice Orthogonality* entails

$$
\begin{aligned}
E(\varepsilon_i &\mid Z_i = 1, C_i = 1) = \\
E(\varepsilon_i &\mid S(Z_i = 0) < s_{\min})) = \\
E(\varepsilon_i &\mid Z_i = 1, C_i = 0) = \\
E(\varepsilon_i &\mid S(Z_i = 0) \geq s_{\min})) = 0
\end{aligned} \tag{11}
$$

while *Weighted Opposite Choice* implies

$$
\frac{E(\varepsilon_i \mid S(Z_i = 0) < s_{\min}))}{E(\varepsilon_i \mid S(Z_i = 0) \geq s_{\min}))} = -\frac{\Pr S(Z_i = 0) \geq s_{\min}}{\Pr S(Z_i = 0) < s_{\min}} = -\Psi_{0,1} \tag{12}
$$

In appendix, we provide two simple illustrations (one incentive-based policy and one dictatorial intervention) using a bivariate normal distribution.

5 Practical Implications

5.1 Policy Design and Instrument Selection

For expositional purposes, assume that an econometrician is allowed to design his/her own policy so to generate a specific instrument.[18] If so, our characterization provides substantial information. For instance, suppose the econometrician is allowed to choose the minimum consumption level. The relevant odd's ratio being inferred from examining the ex-ante consumption

---

[18]Although this may sound awkward, designing the optimal policy for identification purpouses is not much different from experimental protocoles used in Experimental Economics to identify some deep structural (preference) parameters. In the case where a specific parameter raises much interest, say the return to schooling, there is no reason to believe that such an exercise would not be warranted.

distribution, designing a policy requires to span all possible (realistic) values of $s_{\min}$, and evaluate at which level the choice ratio is likely to match the odds ratio.

In the applied literature, an econometrician is sometimes faced with the possibility to choose among a limited set of instruments. Given some priors about the identity of those affected by a policy and the *Noncomplying Odds Ratio*, the econometrician can rely on our characterization to evaluate the reliability of an instrument. In some cases, institutional features or common sense may rule out an extremely high (or low) choice ratio. If so, an instrument may simply be ruled out because of its associated odds ratio.

For instance, an econometrician estimating the return to schooling can use prior information about the degree of compression of the earnings distribution in a particular country, and may decide to rule-out a specific instrument that would require an unrealistic post-schooling earnings skill accumulation ratio.

In the Health Economics example mentioned in Section 3 (in which one is interested in estimating the effect of pollution exposure on birth weight), suppose the econometrician considers using an instrument generated by a policy affecting a sub-population of pregnant women mostly coming from disadvantaged background. Such an instrument could be ruled out if identification requires an unrealistically high average alcohol consumption level for those affected or unaffected. Such a decision would be much more difficult to take on the sole basis of the classical control-treatment identifying condition. Our claim is that recognizing that either *Choice Orthogonality* or *Weighted Opposite Choices* must be fulfilled, allows the econometrician to gain substantial insight about the validity of a specific instrument.

5.2   Implications for Weak Instruments

Our analysis has also implications for the weak-strong instrument literature (Staiger and Stock, 1997). In the literature, a high correlation between the instrument and the right-hand side endogenous variable is viewed as desirable as it increases precision of the IV estimator.

Within our framework, a weak instrument is one that emerges from the implementation of a policy that leaves a large fraction of people unaffected, and which is characterized by a high *Noncomplying Odds Ratio*.[19] As already

---

[19]To be precise, the strength of an instrument is usually measured using the correlation

noted, the *Noncomplying Odds Ratio* plays no role when *Choice Orthogonality* is met. However, when the complying status discloses some degree of selectivity, the fulfillment of the IV identifying condition is itself affected by the statistical strength of the instrument. So, a weak instrument meets the moment condition if it generates an *Error Term Ratio* of the same magnitude as the *Noncomplying Odds Ratio.* Put differently, weak instruments can achieve identification only if they imply important imbalances between compliers and noncompliers unobserved behavior.

This point is different than the usual claim that stronger instruments also reduce the asymptotic IV bias. For this argument to hold, it is necessary to assume that the numerator of the asymptotic bias (the covariance between the instrument and the error term) increases less than the denominator (the covariance between the instrument and the right-hand side endogenous variable) when one moves from one instrument to a stronger one. Indeed, in a context where the error term is made of unobserved choices (such as in the current paper), the numerator is not policy invariant and it is incorrect to claim that stronger instruments automatically reduce the IV asymptotic bias.[20]

5.3   Extensions to More General Outcome Models

In economics, many applications of IV estimation have an inherent life-cycle component and are subject to various forms of dynamics. If choices are made sequentially and consumption of W depends on the time at which it is realized, or if the demand for W depends on consumption level of S (after conditioning on θ_i), the consumption level of good W will also depend on complying status C_i, even after conditioning on θ_i and on all other environmental parameters.[21] This creates no additional complication as long as the econometrician wants to estimate β.

---

between the instrument and the endogenous right-hand side variable. However, it is fair to assume that a large *Noncomplying Odds Ratio* is indicative of a weak instrument.

[20]Belzil and Hansen (2011) illustrates this point within a calibrated life-cycle model of schooling and earnings.

[21]One of the most striking examples is found in the returns to education literature. At any given point in time, the set of individuals affected by an education policy aimed at increasing schooling will suffer a reduction in potential experience, and are likely to have accumulated less skills.

Under certain conditions, our result may even be extended to the case where the econometrician wants to identify the sum of β plus the compound effect of S on W.[22] To remain within a linear IV framework, if one assumes that the effect of S on W is linearly separable,

$$W_i = W_S(S) + W_W(\theta_i))$$

where $W_S(S)$ is linear in S, then it is easy to see that either (7) or (10) must apply to $W_W(\theta_i)$. However, in such a case, the only difference with the stylized example that has been used as the basis of our analysis, is that *Post-Treatment Identity Randomization* will no longer be sufficient for *Choice Orthogonality.* This is because θ_i is no longer sufficient to characterize consumption of good W.[23]

6 Conclusion

The existence of a relationship between IV estimation and sample selection models has been recognized for more than 10 years. In this paper, we show that the relationship is actually more profound than what has been noted in the literature on comparative advantages (and in the literature on the LATE approach). The fundamental IV identifying assumption is itself the mirror image of two mutually exlusive sample selection conditions. These conditions have been derived in a common slope regression model, although they are also valid in a heterogenous treatment effect model.

The originality of our result lies in its constructive nature. Although the interpretation of the IV identifying moment condition as behavioral assumptions has already been pointed out (Rosenzweig and Wolpin, 2000, Keane, 2010), its correspondence with classical sample selection arguments is novel. Under certain conditions, our approach offers more explicit criteria for evaluating the potential performance of an instrument than the usual treatment-control condition, and for understanding the implicit assumptions. Our approach also discloses a link between identification and the weak-strong instrument paradigm. At a more methodological level, our finding should reinforce

---

[22]This is the case in the dynamic treatment effect literature. See Abbring and Heckman (2007) for a survey.

[23]See Belzil and Hansen (2011) for several calibrated examples.

the idea that IV identifying conditions are disguised functional (distributional) form assumptions.

References

[1] Abbring, Jaap and Heckman, James (2007) "Econometric Evaluation of Social Programs, Part 3: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation", *Handbook of Econometrics, Volume 6B.* Elsevier B.V.

[2] Belzil, Christian (2007) "The Return to Schooling in Structural Dynamic Models: A Survey of the Literature" *The European Economic Review,* vol. 51, no 5, 1059-1105

[3] Belzil, Christian and Jörgen Hansen (2011) "The Distinction between Dictatorial and Incentive Policy Interventions and its Implications for IV Estimation" revised version of IZA Working Paper 4835.

[4] Heckman, james (1979) "Sample Selection Bias as a Specification Error," *Econometrica* 47:1 (1979), 153–162

[5] Heckman, James (1997) "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources, 32 (3), 441-62.*

[6] Heckman, James and Vytlacil, Edward (2005) "Structural Equations, Treatment Effects and Econometric Policy Evaluations" *Econometrica. 73.*

[7] Heckman, James, Sergio Urzua and Vytlacil, Edward (2007) "Understanding Instrumental Variables in Models with Essential Heterogeneity", *Review of Economics and Statistics*

[8] Imbens, Guido and Angrist, Joshua (1994) "Identification and Estimation of Local Average Treatment Effects" *Econometrica, 62, 467-76.*

[9] Imbens, Guido (2009) "Better Late than Nothing" , NBER Working Paper

[10] Keane, Michael (2010) "Structural vs. Atheoretic Approaches to Econometrics, *Journal of Econometrics, 156, 3-20.*

[11] Rosenzweig Mark and K.Wolpin (2000) "Natural Natural Experiments in Economics" *Journal of Economic Literature*, December, 827-74.

[12] Staiger, Douglas and James H. Stock (1997),"Instrumental Variables Regression with Weak Instruments" *Econometrica, 65*: 557-586.

[13] Vytlacil, Edward (2002) "Independence, Monotonicity, and Latent Index Models: An Equivalence Result", *Econometrica, 70(1): 331-341*

Appendix
An Example with Normality

Consider a simplistic case in which S is binary ($S = 1$ or $S = 0$). The model generating S is given by

$$S_i = I(\delta + \theta_i > 0)$$

where $I(.)$ is the identity function, δ is a parameter, and where θ is the heterogeneity term.

For simplicity, assume that $\varepsilon_i \sim N(\mu \cdot Z_i, \sigma_\varepsilon^2)$, $\theta_i \sim N(0,1)$. The correlation between ε_i and θ_i is denoted ρ. The moment condition is obviously met is the specific case where $\mu = 0$.

We now consider two different instruments. The first one is a consumption subsidy which increases the utility additively by an strictly positive amount s. The second instrument is dictatorial and generated when the minimum consumption level is set to $s_{\min} = 1$.

Consumption Subsidy

To illustrate the moment condition in the case of a subsidy, it is informative to rewrite the model as

$$S_i = I(\delta + s + \theta_i > 0)$$

It is easy to see that in such a case,

$$\Pr(Z_i = 1, C_i = 1) = \Phi(-\delta) - \Phi(-\delta - s) \tag{13}$$

where $\Phi(.)$ denotes the standard normal cumulative distribution function, and that

$$\Pr(Z_i = 1, C_i = 0) = 1 - \Phi(-\delta) + \Phi(-\delta - s) \tag{14}$$

The non-complying odds ratio is therefore

$$\Psi_{0,1} = \frac{1 - \Phi(-\delta) + \Phi(-\delta - s)}{\Phi(-\delta) - \Phi(-\delta - s)}$$

Using the properties of the bivariate Normal distribution, we obtain the following:

$$E(\varepsilon_i \mid Z_i = 1, C_i = 1) = \mu + \sigma_\varepsilon \rho \{ \frac{\phi(-s-\delta) - \phi(-\delta)}{\Phi(-\delta) - \Phi(-\delta - s)} \} \qquad (15)$$

and

$$\begin{aligned} E(\varepsilon_i \mid \;\; & Z_i = 1, C_i = 0) = \\ & \frac{(1 - \Phi(-\delta))}{1 - \Phi(-\delta) + \Phi(-\delta - s)} \cdot (\mu + \sigma_\varepsilon \rho \frac{\phi(-\delta)}{1 - \Phi(-\delta)}) + \\ & \frac{\Phi(-\delta - s)}{1 - \Phi(-\delta) + \Phi(-\delta - s)} \cdot (\mu - \sigma_\varepsilon \rho \frac{\phi(-\delta - s)}{\Phi(-\delta - s)}) \qquad (16) \end{aligned}$$

So, the moment condition is either verified when $(15) = (16)$ or when the ratio of (15) over (16) is equal to $\frac{1 - \Phi(-\delta) + \Phi(-\delta - s)}{\Phi(-\delta) - \Phi(-\delta - s)}$.

Minimum Consumption

When the minimum consumption level is set to $s_{\min} = 1$, we obtain

$$\Pr(Z_i = 1, C_i = 1) = \Phi(-\delta) \qquad (17)$$

and

$$\Pr(Z_i = 1, C_i = 0) = 1 - \Phi(-\delta) \qquad (18)$$

The non-complying odds ratio is therefore

$$\Psi_{0,1} = \frac{1 - \Phi(-\delta)}{\Phi(-\delta)}$$

Because consumption is monotonic in θ_i, the intervention splits the population into two sub-populations which may be characterized in terms of a single reservation value $(-\delta)$, and it is clear that a minimum consumption intervention cannot fulfill *Choice Orthogonality*. To see this, we just note that

$$E(\varepsilon_i \mid Z_i = 1, C_i = 1) = \mu - \sigma_\varepsilon \rho \frac{\phi(-\delta)}{\Phi(-\delta)} \qquad (19)$$

and

$$E(\varepsilon_i \mid Z_i = 1, C_i = 0) = \mu + \sigma_\varepsilon \rho \frac{\phi(-\delta)}{1 - \Phi(-\delta)} \tag{20}$$

However, *Weighted Opposite Choices* is verified if

$$\frac{\mu - \sigma_\varepsilon \rho \frac{\phi(-\delta)}{\Phi(-\delta)}}{\mu + \sigma_\varepsilon \rho \frac{\phi(-\delta)}{1-\Phi(-\delta)}} = \frac{1 - \Phi(-\delta)}{\Phi(-\delta)} \tag{21}$$