Identification with possibly invalid IVs

Christophe Bruneel-Zupanc*
Department of Economics, KU Leuven

Jad Beyhum†
Department of Economics, KU Leuven

January 8, 2024

Abstract

This paper proposes a novel identification strategy relying on *quasi-instrumental variables* (quasi-IVs). A quasi-IV is a relevant but possibly invalid IV because it is not completely exogenous and/or excluded. We show that a variety of models with discrete or continuous endogenous treatment, which are usually identified with an IV - quantile models with rank invariance additive models with homogenous treatment effects, and local average treatment effect models - can be identified under the joint relevance of two complementary quasi-IVs instead. To achieve identification we complement one *excluded* but possibly endogenous quasi-IV (e.g., "relevant proxies" such as previous treatment choice) with one *exogenous* (conditional on the excluded quasi-IV) but possibly included quasi-IV (e.g., random assignment or exogenous market shocks). In practice, our identification strategy should be attractive since complementary quasi-IVs should be easier to find than standard IVs. Our approach also holds if any of the two quasi-IVs turns out to be a valid IV.

Keywords: instrumental variables, identification, nonseparable models, selection models, treatment effects, exclusion restriction.

*E-mail address: christophe.bruneel@gmail.com

†E-mail address: jad.beyhum@gmail.com. We are especially grateful to Geert Dhaene for his extensive comments. We also thank Laurens Cherchye, Xavier d'Haultfoeuille, Alexandre Gaillard, Yagan Hazard, Thierry Magnac, Martin Mugnier, Lorenzo Navarini, Florian Schuett, Frank Verboven, and Frederic Vermeulen for useful comments. We acknowledge financial support from the KU Leuven Research Funds through the grants STG/21/040 and STG/23/014.

1 Introduction

A popular quest in empirical economics is to recover the causal effect of a treatment variable D (e.g., education), on an outcome variable Y (e.g., earnings). With nonexperimental data, D and unobserved heterogeneity U (e.g., ability) may not be independent. In this case, causal effects are not characterized by the conditional distribution of Y given D, and D is deemed endogenous. A convenient strategy to address endogenous selection is to rely on instrumental variables (IVs). A valid IV must be relevant for the selection and "strongly excluded" , i.e., be independent of the unobservables U (exogeneity) and have no direct effect on the outcome (exclusion). In practice, however, instruments satisfying these conditions are hard to find. In particular, the exclusion and exogeneity restrictions are rarely jointly satisfied and are often controversial, even for commonly used instruments.

In practice, many variables are not completely valid IVs, but are nonetheless relevant and satisfy either exogeneity or exclusion but not both: we call these "*quasi-instrumental variables*" (quasi-IVs). In this paper, we propose a novel identification strategy relying on two complementary quasi-IVs, Z and W, which are invalid IVs on their own but "complementarily valid" taken as a set. The first variable, Z, is an *excluded quasi-IV* that is possibly endogenous with respect to the unobservables, U, but is relevant (even when controlling for U) and has no direct effect on Y, apart from its indirect effect through D, U, and W. The second variable, W, is exogenous *conditional on* Z and relevant but possibly included. We call it the *exogenous quasi-IV*. In addition, complementary validity requires that these two quasi-IVs are jointly relevant for the selection into treatment, i.e., ceteris paribus (conditional on U), the effect of Z on D varies with W (or vice versa). Separate relevance of W and Z is a necessary condition for the joint relevance.

In practice, finding two complementary quasi-IVs should be easier than finding one valid IV, since the separate requirements on each of the variable are weaker than exogeneity and exclusion combined.

If either of the two quasi-IVs turns out to be a valid IV, our identification strategy is still valid. To sum up, we relax the joint exclusion and exogeneity of IVs at the cost of a stronger joint relevance of W and Z for the selection into D.

One can think of many examples of such complementary exogenous and excluded quasi-IVs. The excluded quasi-IV, Z, can be thought of as a "relevant proxy": a proxy for the unobservable, U, that is also directly relevant for the treatment, even when controlling for U, but whose effect on Y only goes through U and D. Many variables fit this description. For instance, past school grades or schooling recommendations by teachers in the schooling example, or more generally past treatments whose effect on the outcome is superseded by the current treatment in dynamic models with adjustment costs. To complement the excluded quasi-IV, one needs to find a complementary exogenous quasi-IV, i.e., a variable which is independent from U conditional on Z. Conditional exogeneity differs from exogeneity and could be violated even with unconditionally exogenous W, if Z was determined as a function of W for example. In practice, a convenient way to satisfy this conditional independence is to find a variable which satisfies the even stronger requirement of being jointly independent from U and Z by exploiting the timing of the realization of the variables. Typically, unanticipated exogenous shocks to local markets or local policy changes should be exogenous with respect to individuals' observed and unobserved characteristics but still affect individuals' outcomes directly. The timing of Z occuring before W guarantees that Z could not depend on W if W was unanticipated.

Random assignment in randomized experiments are other natural examples of exogenous quasi-IVs (e.g. Bloom et al., 1997; Heckman et al., 1997, 1998; Abadie et al., 2002; Schochet et al., 2008, ...). The randomization guarantees its exogeneity with respect to any pre-treatment variables and that it could not be anticipated beforehand. However, these assignments are not necessarily valid IVs because there are often reasonable concerns that they may be included and have a direct effect on the outcome, even after controlling for the treatment. For example, winning a randomly assigned voucher to cover the cost of private school has a direct effect on the subsequent educational outcomes (Angrist et al., 2002). There is an included income effect because the family of winners who attend private school are richer, and moreover, the lottery assignment directly affects the level of effort exerted by the students. Similarly, the military draft (Angrist, 1990) may have a direct effect on drafted individuals because they may behave differently to avoid going to war (by getting more education) and conscientious objectors may go to prison because they did not comply.

Another crucial example is (fuzzy) Difference-in-Differences (DiD), see Card and Krueger (1994); Athey and Imbens (2006); De Chaisemartin and d'Haultfoeuille (2018). In DiD, the group assignment corresponds to our Z, it is related to unobserved characteristics but typically excluded from the model given these characteristics. Time plays the role of W, it is exogenous (conditional on the group assignment) but is allowed to have a direct effect on outcomes. Moreover, W and Z (time and group) are jointly relevant for the selection into treatment, to the point where, in the sharp design, the interaction $W \times Z$ is the treatment.

From a theoretical perspective, we show that the models which are usually nonparametrically identified with an IV can be nonparametrically identified with two complementary quasi-IVs instead. In particular, we prove nonparametric identification of quantile models with rank invariance (Chernozhukov and Hansen, 2005), additive models with homogenous treatment effects (Newey and Powell, 2003; Darolles et al., 2011), and local average treatment effect (LATE) models (Imbens and Angrist, 1994; Heckman and Vytlacil, 2005). The intuition behind our result is that the complementary validity of our two quasi-IVs implies that they only have a separable effect on the outcome but a "jointly relevant" nonseparable effect on the selection into treatment.[1] In this case, the joint relevance on D provides exogenous variation that is excluded from Y and that we can generally exploit to achieve identification. In the special case of a linear additive model, the complementarity of the two quasi-IVs is equivalent to positing that the interaction between the two quasi-IVs is a valid IV (see the discussion at the end of Section 3). Indeed, the interaction is relevant (by joint relevance) but excluded from the outcome equation (by separability of the effect of the two quasi-IVs on the outcome). In a more general nonlinear context, the complementarity provides a general form of exclusion restriction via the separability of the effects of the two quasi-IVs. In our main framework with one excluded (Z) and one conditionally exogenous (W) quasi-IV, W and Z only have a separable effect on Y by construction. Indeed Z is excluded from Y if we control for the unobservables (and for D), and W is excluded from these unobservables by exogeneity.

[1] Concerning the LATE framework, we allow the exogenous quasi-IV W to be possibly included with an homogenous effect (this is similar to the fact that time is typically assumed to have an homogenous effect in the difference-in-differences literature). We also need a testable local irrelevance assumption positing that there exists a level of the excluded IV such that the exogenous IV does not affect the propensity score.

Related literature. This paper is related to a strand of the econometric literature identifying econometric models with endogeneity without (explicitly) relying on instrumental variables. These papers address endogeneity by imposing parametric functional form restrictions (see Rigobon, 2003; Dong, 2010; Klein and Vella, 2010; Lewbel, 2012; Escanciano et al., 2016; Lewbel, 2018; D'Haultfœuille et al., 2021; Jiang and Tsyawo, 2022; Tsyawo, 2023; Lewbel et al., 2023; Gao and Wang, 2023). These restrictions serve to confine the structural regression function's degrees of freedom, enabling the identification of the causal effect of the endogenous variable through nonlinear variations induced by exogenous (yet included) variables. Our approach, grounded in separability, can be viewed as a nonparametric counterpart to such parametric conditions. We also give a specific economic justification to the separability by introducing the concept of complementary quasi-IVs.

We note that separability-type restrictions have also been used in settings with standard (excluded and exogenous) instruments to obtain identification when the relevance condition may not hold, see Caetano and Escanciano (2021); Feng (2024) among others. Here, we rather use separability to address lack of exogeneity or exclusion. From a general perspective, separability is a common tool in econometrics. For instance, the separability of fixed effects with respect to time and subjects is what allows identification of panel data models with two-way fixed effects (including difference-in-differences models).

Another line of works studies identification of causal effects using exogenous but included instrumental variables. Liu et al. (2020) point identifies the partial derivative of the average treatment effect with respect to an exogenous quasi-IV, but they need this included instrument to be excluded from the selection equation. D'Haultfœuille et al. (2021) adopts a control function framework using an exogenous quasi-IV, achieving point identification through a local irrelevance condition. We employ a similar restriction in identifying our LATE model. Wang (2023) point identifies the LATE with an exogenous quasi-IV satisfying the standard monotonicity assumption of Imbens and Angrist (1994) and another exogenous and excluded IV that can violate monotonicity. Some other works bound the treatment effects with an exogenous quasi-IV IV, see Manski and Pepper (2000); Nevo and Rosen (2012); Conley et al.

(2012); Flores and Flores-Lagunes (2013); Mealli and Pacini (2013); Ban and Kédagni (2022) among others. Bartels (1991) also uses the term "quasi-IV", but referring to variables that can be both included and endogenous, but only slightly deviating from the exclusion/exogeneity requirements. Bartels (1991) then conducts a sensitivity analysis in this setting. In our case, the suffix "quasi" suffix means that the variables only satisfy part of the IV requirements, but in the dimension from which they deviate from the standard, the violations of the exclusion or exogeneity can be very large.

The present paper is also related to Bruneel-Zupanc (2023), which introduces semi-instrumental variables (semi-IVs) in the context of discrete treatment variables. A semi-IV is an exogenous variable that is only excluded from some (but not all) potential outcomes. Identification with semi-IVs works for similar reasons as identification with complementary quasi-IVs: it requires complementary semi-IVs, i.e., at least one semi-IV excluded from each potential outcome (hence a complementarity in the exclusion), satisfying a stronger joint relevance condition than standard IVs. The two papers are related but study identification with a completely different type of invalid IVs. The advantage here is that quasi-IVs also apply to model with continuous endogenous variables while semi-IVs cannot by construction (because one would need to find infinitely many of semi-IVs for infinitely many potential outcomes). The assumptions, results and proofs of the two papers differ.

A large set of relevant empirical work also employs variables sharing similarities with our exogenous variables (W) and excluded quasi-instrumental variables (Z). The literature on Bartik/Shift-share instruments typically hinges on several exogenous shocks to which the units have (known) different level of (endogenous) exposures (Bartik, 1991; Adao et al., 2019; Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022; Borusyak and Hull, 2023). In this case, they have that the exogenous shocks are independent from the unobservables, even conditional on the (possibly endogenous) exposures. So the conditional exogeneity holds. However, for identification they use a "known formula" (formula instrument) interacting the exogenous shocks with the nonrandom exposures to create the shift-share IV. The manner we achieve identification is quite different. The first difference is, we focus more directly on local shocks as W since we need variations in W within sample. In a sense our W could be the shift-share IV directly. But we allow for this W to be included and have a direct effect on the outcome, even controlling for D and U. Second, endogenous exposures are probably not valid excluded quasi-IVs for our approach because they may not be excluded. Third, we do not rely on a known formula to build an instrument, our procedure is valid with general Z, not only exposures or related variables, for which no known formula is available, it just requires Z to be a credible excluded quasi-IV. The manner in which W and Z interact to impact D is completely unknown and free: we do not make any parametric assumption and allow the relation to be completely nonseparable, even with respect to the unobservables, U.

Finally, as previously noted, in the context of difference-in-differences (Card and Krueger, 1994; Athey and Imbens, 2006; De Chaisemartin and d'Haultfoeuille, 2018), time serves the role of our variable W, and Z corresponds to the group assignment. Let us compare our results with that of the literature on DiD. In quantile models with rank invariance, we can identify quantile treatment effects over the whole population thanks to a rank invariance assumption with respect to both treatment and time. Our framework allows the treatment and time to be continuous and a possibly infinite number of groups. As a comparison, in sharp DiD designs, Athey and Imbens (2006), only imposes rank invariance with respect to time (but not to treatment) and they can only identify treatment effects on the treated or untreated groups, but this relies crucially on the fact that the design is sharp (a restriction that we avoid). In fuzzy DiD designs, De Chaisemartin and d'Haultfoeuille (2018) also make a type of rank invariance assumption with respect to time, and identify effects on the compliers. In this fuzzy setting, we can identify effects over the whole population at the cost of assuming rank invariance with respect to the treatment as well. Our findings on the LATE can be viewed as an extension of these previously mentioned results from De Chaisemartin and d'Haultfoeuille (2018) on fuzzy DiD, as we allow for continuous time and an infinite number of groups.

Outline. The paper is organized as follows. In Section 2, we show identification in a quantile model with rank invariance and discuss further concrete examples of the quasi-IVs. Then, Section 3 contains identification results for an additive model with homogenous treatment effects. The LATE framework is discussed in Section 4.

Section 5 concludes. The proofs of our formal results can be found in the appendix.

2 Quantile model with rank invariance

2.1 The model

In this section, we consider identification in a quantile model with rank invariance. Let D, with support $\mathcal{D}$, be the observed (endogenous) choice/treatment and Y_d the (continuous) latent potential outcomes under treatment state d, for $d \in \mathcal{D}$. The potential outcomes $\{Y_d\}$ are latent because we only observe one outcome, $Y = Y_D$, corresponding to the potential outcome of the selected alternative D. Let Z be an excluded (possibly endogenous) quasi-IV, whose possible effect on Y only goes through D, W, X and the unobservables U. Denote by W the complementary exogenous quasi-IV (conditional on Z), which may be "included" and have a direct effect on Y. By construction, the effects of W and Z on Y are separable, in the sense that, if we could control for U (in addition of D and W), Z would have no effect on Y. We also allow for additional covariables X. The supports of W, Z, X are denoted by $\mathcal{W}, \mathcal{Z}, \mathcal{X}$, respectively. We study the following structural quantile regression model with endogeneity:

$$Y_d = f(d, W, X, U),\ U \sim \mathcal{U}[0,1], \tag{1}$$

for all $d \in \mathcal{D}$, where f is strictly increasing in its last argument. The treatment is possibly endogenous since we do not assume that $U \perp\!\!\!\perp D|(W, X)$. Z is excluded from the potential outcomes given W and U (and covariables X).

To address the endogeneity issue, we assume that there exists a mapping $g : \mathcal{Z} \times \mathcal{X} \times [0,1] \to \mathbb{R}$ strictly increasing in its last argument such that

$$U = g(Z, X, V),\ V|Z, W, X \sim \mathcal{U}[0,1], \tag{2}$$

Equation (2) is equivalent to key identification condition

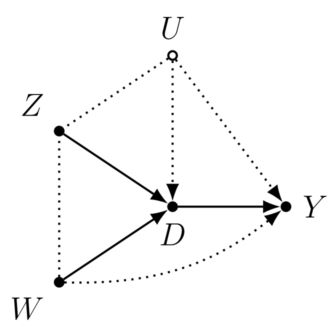
$$U \perp\!\!\!\perp W|(Z, X),$$

that is W is exogenous given Z (and covariables X). However, W is not excluded from (1) and Z can be endogenous through (2). If they also satisfy a joint relevance condition, W and Z are "complementarily valid" quasi-IVs in this Framework. The general covariables X are not quasi-IVs, they differ from W or Z because they are not excluded from neither (1) nor (2). The facts that U and V have a uniform distribution are just some normalizations. Note that, for $w \in \mathcal{W}, x \in \mathcal{X}, z \in \mathcal{Z}$ and $v \in [0,1]$, $f(d,w,x,g(z,x,v))$ is the v-quantile of the distribution of Y_d given $W = w, X = x, Z = z$. The model in (1) implies rank invariance (see Chernozhukov and Hansen, 2005). It essentially means that rank in the outcome of any two subjects with the same value of W, Z, X is the same across all potential outcomes.[2]

Importantly, notice that our framework allows W to be excluded and/or Z to be exogenous. It nests the case where any (or both) of the two quasi-IV turns out to be a valid IV.

Our framework can be visualized through the causal graph of Figure 1. We do not represent X here to simplify the graph. The fact that the only link between W and U passes by Z means that W and U are independent given Z, ensuring conditional exogeneity of W. There is no direct arrow from Z to Y, so that Z is excluded given D, W and U. W and Z can be correlated, even though in most application they will not be. We caution the reader that causal graphs are only a visualization tool but do not rigorously encode all probabilistic assumptions.

Figure 1: Causal graph of the Framework



Solid and empty nodes represent observed and unobserved variables, respectively. Solid arrows indicate causal effects. Dotted arrows indicate possible causal relationships. Dotted undirected lines indicate possible general relationships without specifying the direction.

Alternative potential outcome justification.

[2]We could relax this assumption and instead impose rank similarity as in Chernozhukov and Hansen (2005), but we do not do so to avoid complicating the exposition.

In the baseline Framework described above, U can be seen as a structural error term that has a clear structural interpretation (e.g., unobserved ability, preference, productivity, etc.). Alternatively, we could write the problem in terms of potential outcomes with

$$Y_d = \tilde{f}(d, W, X, \tilde{U}_d), \ \tilde{U}_d|(W, X) \sim \mathcal{U}[0, 1], \tag{3}$$

There, $\tilde{U}_d = F_{Y_d|W,X}(Y_d|W, X)$ is the rank of Y_d in its distribution given W and X, and by construction, $\tilde{U}_d \perp\!\!\!\perp (W, X)$. Then, the rank invariance, means that there exists a variable $\tilde{U}$ such that $\tilde{U}_d = \tilde{U}$ for all d. This variable may depend on Z, and we can write it as

$$\tilde{U} = \tilde{g}(Z, X, V), \ V|Z, W, X \sim \mathcal{U}[0, 1]. \tag{4}$$

This implies our main assumption (the complementarity), that is $\tilde{U} \perp\!\!\!\perp W|Z, X$. Very importantly, notice that $\tilde{U} \perp\!\!\!\perp W|X$, which is guaranteed by construction, is a completely different assumption from $\tilde{U} \perp\!\!\!\perp W|Z, X$. To satisfy $\tilde{U} \perp\!\!\!\perp W|Z, X$, a necessary condition is that Z has a separable effect from W on Y_d.[3] The baseline Framework actually fits into this framework written in terms of potential outcomes, and would yield $\tilde{U} \perp\!\!\!\perp W|Z, X$ (even though U is generally different from $\tilde{U}$). The two frameworks are equivalent when it comes to the identification arguments.

2.2 Discussion and structured examples

Let us discuss the conditions under which W and Z are complementarily valid quasi-IVs and provide several examples of applications. Every statements are implicitly conditional on X and we only mention X when necessary.

The difference between $\mathbf{W} \perp\!\!\!\perp \mathbf{U}$ and $\mathbf{W} \perp\!\!\!\perp \mathbf{U}|\mathbf{Z}$.

Our framework requires the conditional independence that $W \perp\!\!\!\perp U|Z$. This is different from unconditional exogeneity, $W \perp\!\!\!\perp U$. Finding an exogenous W is not sufficient (nor necessary), one needs to find a W that is exogenous when conditioning on Z. In particular, we may run into a problem if Z is a function of W, i.e., $Z = \varphi(W, X, U)$. In this case, even if W is unconditionally exogenous, i.e., $W \perp\!\!\!\perp U$, we will not gener-

[3] Note again that X is neither exogenous nor excluded. Even if $\tilde{U} \perp\!\!\!\perp X$, we may not have $\tilde{U} \perp\!\!\!\perp X|Z$, which is the reason why X must appear inside the function $\tilde{g}$ in (4).

ally have $W \perp\!\!\!\perp U|Z$.

In practice, thinking in terms of conditional independence is often nontrivial, therefore, to find valid W and Z, we suggest to look for variables satisfying the stronger, but more intuitive, requirement that $W \perp\!\!\!\perp (U, Z)$.

To ensure this stronger condition when Z is a choice, we need that W is excluded from the decision Z. This can be achieved by exploiting the timing of the variables, if W is unknown and unanticipated when Z is decided. For example, in a dynamic model where Z is a choice taken in period $t-1$, our assumption hold if W is an exogenous shock occuring in period t that was unanticipated in period $t-1$. This is why, in many examples below, we use “past decisions/outcomes” as Z, and exogenous shocks (i.e. contemporaneous variations) to the local market conditions as W. We do not use local market conditions themselves as W because these may partly affect Z.

Nonetheless, it is also possible that $W \perp\!\!\!\perp U|Z$ even if $W \not\perp\!\!\!\perp U$. A typical example of this is when W represents randomized treatment but the randomization occurs on the basis of a specific variable, e.g., unemployment status, education or income. Then, taking these variables as Z, we have $W \perp\!\!\!\perp U|Z$, and the requirement is then that the Z are plausibly excluded conditional on D, U and W. Therefore, our work implies that randomized experiments can be “conditionally randomized” on the basis of observable excluded quasi-IVs to favor more some target population (e.g., the unemployed or uneducated), and still identify treatment effects for everyone. If one expects positive effects from the treatments, being able to identify the treatment effect with this conditional randomization is a desirable property, especially when the program is costly.

Example 1 (Returns to schooling). We are interested in the effect of schooling (years of education, or binary decision to go to college or not), D, on the earnings at age 30, Y (Card, 1995, 2001). We have an endogeneity issue because the decision to go to college and the subsequent earnings are both impacted by unobserved ability, U. To identify the effect of D on Y, we need to find two complementary quasi-IVs. Let us consider past grades or school recommendation by teachers as the excluded quasi-IV, Z. The idea is that these variables are "relevant proxies" in the sense that, they are likely strongly correlated with individuals' unobserved abilities (U), but once you control for U and D (and X), they have no effect on the outcome. Moreover, they are still relevant, even controlling for U, because everything else equal (i.e., at fixed ability), a student with better grades or with a good recommendation, is more likely to go to college. We complement these excluded IVs by using exogenous shocks to the local market conditions (wage levels, unemployment rates, tuition fees, ...) at the time of the decision to go to college (around age 17) as the exogenous quasi-IV, W. These types of exogenous shocks were used as IVs by Carneiro et al. (2011) when controlling for permanent market conditions for example. However, these W may not be valid IVs because they may be included if, for example, a positive shock to local market condition at age 17 persists over time. Yet, they are valid W because they are likely exogenous with respect to unobserved student ability, U, and past individuals' grades/school recommendations, Z, so $W \perp\!\!\!\perp (U, Z)$, which is stronger than the required $W \perp\!\!\!\perp U|Z$. Notice that it is important that W was unanticipated at the time when the student grades were obtained, otherwise, some students may have potentially worked harder because of W, and the conditional independence of $W \perp\!\!\!\perp U|Z$ would not necessarily hold then. The two complementary quasi-IVs are also likely to be jointly relevant if the likelihood of pursuing education for students with good grades compared to student with bad grades varies with market conditions.

Example 2 (Returns to private schooling using randomized lotteries). Another typical question is the returns to private vs. public schooling (D) on educational attainment, Y. Again, the main problem is that both the educational attainment and the decision to go to a private school are endogenous with respect to the unobserved student ability. The Colombian government ran a large scale randomized experiment, the PACES program, giving vouchers which partially covered the cost of private secondary school (Angrist et al., 2002). The vouchers were assigned randomly via a lottery, so their assignment is exogenous by construction, and it is a natural candidate exogenous quasi-IV, W. However, despite its exogeneity, W is probably not a valid IV if we want to study the returns to private schooling, because the vouchers may be included and have a direct effect on educational attainment. For example, conditional on going to private schooling, families of students receiving the vouchers are richer than families who did not receive it. This income effect may directly affect the subsequent educational attainment of the children. Another concern is that student with vouchers may exert a different amount of effort than students without vouchers to get into the same schools. Because of these concerns, Angrist et al. (2002) study the effect of winning the lottery on educational attainment, not the effect of private schooling. But using our methodology, we can actually leverage the exogeneity of W to identify the causal effect of private schooling. To do so, we need to find an excluded quasi-IV, Z, to complement the exogenous voucher assignment, W. For the same reasons as in the previous returns to schooling example, variables such as pre-lottery student grades are valid Z: they are relevant for the selection into private school, even controlling for unobserved ability U, but are otherwise excluded from the subsequent schooling attainments when one controls for D and U. Moreover, because the lottery was randomized, and especially if the PACES program was unanticipated, Z is also independent from the voucher assignment, W. So we have $W \perp\!\!\!\perp (U, Z)$, and thus, $W \perp\!\!\!\perp U|Z$. The joint relevance is likely to be satisfied, as the increase in the probability to attend private school if receiving a voucher varies with the past school achievements, Z.

Example 3 (Marginal returns to hours worked). Suppose we want to identify the marginal effect of working more on earnings, Y. In this case, D represent the (continuous) worked hours, or a general measure of work flexibility. The main endogeneity issue is that the hours worked decision depends on individuals' unobserved productivity, U. In order to address this endogeneity, we need two complementary quasi-IVs. If we focus on a sample of married men, let us use the spouse characteristics (e.g., spouse education or income) as a Z (e.g., Mroz, 1987). Z is correlated with the wife's unobserved productivity, and in the presence of assortative matching, the wife's unobserved productivity should correlate with the husband's unobserved productivity, U. It is also possible that the wife's characteristics have an effect on U directly. As a consequence, Z is a not a valid IV because it is not exogenous. However, Z should be a valid excluded quasi-IV for the following reasons. First, it is likely very relevant for the husband working time decision, even controlling for his unobserved productivity, because it has a direct effect on the household available income. Second, controlling for the unobserved productivity, U, and for the hours worked, D, and for other control variables such as husband's experience, X, the wife's characteristics, Z, should have no direct effect on the husband earnings, Y. Z is thus a plausible "relevant proxy" and thus excluded quasi-IV. A similar reasoning can be applied to many other plausible Z in this application. For example, the simple fact of being married or not, or other variables representing the number of children or their age could be valid excluded IVs, Z: it is plausible that they do not have any effect on Y conditional on D and U. To complement these excluded quasi-IVs, we can use exogenous shocks to the local market characteristics as W. The conditional independence condition, $W \perp\!\!\!\perp U|Z$, is more likely to hold if the decision to marry/to have kids was independent from W. It is plausible if the marriage happened in a previous period while W are contemporaneous market shocks. As for the joint relevance, it is likely to hold as the marginal effect of Z on D should vary with the market shocks W.

Example 4 (Dynamic models with adjustment costs). Let us consider dynamic models with adjustment costs. For example, one can think of a model of labor supply, D_t, and consumption, Y_t to study the marginal propensity to consume with respect to work (Blundell et al., 2016a,b; Bruneel-Zupanc, 2022).[4] There is an endogeneity issue in this case because unobserved individual preference for consumption, U_t, affects both decisions D_t and Y_t. In this case, D_{t-1} is a natural excluded quasi-IV, Z_t: conditional on the current D_t, and controlling for current covariates such as wealth and hourly wage in X_t, the past labor decision, D_{t-1} has no effect on the current consumption choice, Y_t. Moreover, in the presence of adjustment costs in the labor supply, the current labor supply decision is directly affected by the previous labor supply choice, even controlling for U_t and X_t. D_{t-1} is, however, not a valid IV because it is likely correlated with U_t as soon as there is some persistence in the preference for consumption. We complement this excluded quasi-IV with contemporaneous exogenous shocks to the local market conditions (e.g. to prices) which were unpredictable at time $t-1$ and are independent from the individual unobserved preferences. Thus, $W_t \perp\!\!\!\perp (U_t, Z_t)$.

[4] One could also consider the identification of dynamic production functions (Olley and Pakes, 1996; Blundell and Bond, 2000; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020, see De Loecker and Syverson, 2021 for a recent survey) with adjustment costs on the inputs as a similar alternative example.

As stressed earlier, it is important that W_t was unanticipated at $t-1$ when D_{t-1} was decided, because other D_{t-1} would depend on W_t. Finally, joint relevance should be satisfied if the effect of D_{t-1} on D_t varies with the contemporaneous exogenous shocks.

Example 5 (Job Training Programs). We are interested in the effect of job training programs such as JTPA (Bloom et al., 1997; Heckman et al., 1997, 1998; Abadie et al., 2002) or Job Corps (Schochet et al., 2008) on subsequent job market outcomes, Y (employment or earnings). In this case, random assignment into the program is exogenous by construction. However, it is only a valid IV if we want to recover the global effect of the program. Typically, job training programs offer a variety of trainings. Imagine for example that everyone receives a job search assistance, but only some individuals choose to follow a schooling training in addition. If we want to recover the effect of schooling, D, on Y, the random assignment to the program is not a valid IV anymore. Indeed, despite its exogeneity, it is included since assigned individuals also obtain the job search assistance, which may have a direct effect on the outcome. A solution to address this issue is to find a complementary excluded quasi-IV, Z. A good candidate in this setup is pre-assignment schooling achievement. If we consider D as the schooling level after training, Z is excluded from Y conditional on D and U because the subsequent schooling level supersedes the previous one. Previous schooling is however relevant for the selection into the training since, everything else equal, individuals with different schooling background may be more or less likely to accept to participate in the training. It is however not a valid IV since it is strongly correlated with unobserved ability, U. Since Z is a pre-treatment variable and the random assignment was not anticipated, $W \perp\!\!\!\perp (U, Z)$. Moreover, the joint relevance is satisfied since the increased probability of following the treatment if assigned to it should vary with the previous schooling achievement. So, W and Z are complementarily valid quasi-IVs.[5]

[5] Treatment recommendations or measures of long term unemployment probability by the training agencies are also typically good Z. One concern here, is that, when there are several kind of treatments, the recommendations may not be excluded from Y. Indeed, an individual to whom we recommended to follow a school training may follow another non-mandatory training if assigned to the program. As a consequence, assigned individuals ($W = 1$) with schooling recommendation ($Z = 1$) and assigned individuals with another recommendation ($Z = 0$) do not follow the same trainings, and the interaction between W and Z has a direct included effect on the outcome. This is not a concern if the alternative trainings are mandatory and followed by everyone. Otherwise, we

Example 6 (Difference-in-differences). Consider a difference-in-differences setting where W is time and Z is the group assignment. D is typically a binary treatment. Our key assumption that $U \perp\!\!\!\perp W|Z$ is standard in the DiD literature. Indeed, Athey and Imbens (2006) imposes "time invariance within groups" (their Assumption 3.3) and De Chaisemartin and d'Haultfoeuille (2018) assumes "time invariance of unobservables" (see their Assumption 7). It states that time is independent from unobserved heterogeneity within each group or, in other words, that the composition of the groups does not change with time. This is exactly $U \perp\!\!\!\perp W|Z$ in our model. In DiD, this assumption is for instance satisfied by construction if one observes all units in each group at every date.

Compared to Athey and Imbens (2006) and De Chaisemartin and d'Haultfoeuille (2018), we impose rank invariance with respect to both time and treatment. Instead, they only impose rank invariance with respect to time of a given potential outcome (see their Assumptions 3.1 and 7, respectively). However, in fuzzy designs where units in a given group can have different treatment statuses, the approach of these two papers can only identify local quantile treatment effects. The results in the present section instead allow to identify quantile treatment effects over the whole population. We also note that, in DiD papers, treatment, time, and group assignments are usually all discrete. Here, we allow them to be continuous. Note that we continue this analysis of our results in the context of DiD in the LATE framework in Section 4.

2.3 Identification of the model

2.3.1 System of equations

In this section, we study nonparametric identification of different unknown (infinite dimensional) parameters. For simplicity, we will suppress the dependence on X. All our results and assumptions can be interpreted as conditional on X. For $u \in [0,1], z \in \mathcal{Z}, w \in \mathcal{W}$, let $F_{U|Z}(u|z) = \mathbb{P}(U \leq u|Z = z)$.

We want to identify the functions f and g. Notice that, for all $u \in [0,1]$ and $z \in \mathcal{Z}$, we have $F_{U|Z}(u|z) = g(z,\cdot)^{-1}(u)$. Because of this one-to-one relationship between g and $F_{U|Z}$, we focus on identification of $F_{U|Z}$ rather than of g, but the two

need to control for the different kind of trainings (if observable).

are equivalent.

We observe the joint distribution of (Y, D, W, Z). Identification relies on the fact that the model and assumptions imply that, for $u \in [0,1]$, $w \in \mathcal{W}$, and $z \in \mathcal{Z}$, we have

$$\begin{aligned} F_{U|Z}(u|z) = & \quad \mathbb{P}(U \leq u|W = w, Z = z) \\ = & \quad \mathbb{E}[\mathbb{E}[\mathbb{1}(U \leq u)|D, W, Z]|W = w, Z = z] \\ = & \quad \mathbb{E}[\mathbb{E}[\mathbb{1}(Y_D \leq f(D, W, u))|D, W, Z]|W = w, Z = z] \\ = & \quad \mathbb{P}(Y \leq f(D, w, u)|W = w, Z = z). \end{aligned} \tag{5}$$

The first equality comes from $U \perp\!\!\!\perp W|Z$ because $V \perp\!\!\!\perp (Z, W)$. Then, we rewrite the equation using the law of total expectation and the next equality follows from the monotonicity of f in its last argument. The last equality is a consequence of the fact that $Y = Y_D$ and the law of total expectations.

Moreover, for $u \in [0,1]$, we have

$$u = \mathbb{E}[F_{U|Z}(u|Z)], \tag{6}$$

by the law of total expectation. This gives one more equation.[6]

The function f and the conditional distribution functions $F_{U|Z}$ are identified from the joint distribution of (Y, D, W, Z) if they are the unique solution in a given (infinite dimensional) parameter space to the continuum of integral equations in (5)-(6). In the next two sub-subsections, we give relevance conditions guaranteeing the uniqueness of the solution to (5)-(6) when D is discrete (Section 2.3.2) and when D is continuous (Section 2.3.3).

2.3.2 Identification with discrete D

Let us assume in this section that D is discrete with $\mathcal{D} = \{1, ..., |\mathcal{D}|\}$. Without loss of generality, we can also suppose that W and Z have discrete support $\mathcal{W} = \{1, ..., |\mathcal{W}|\}$ and $\mathcal{Z} = \{1, ..., |\mathcal{Z}|\}$. In this setting, (5)-(6) become

[6]There are more equations of the form $\mathbb{P}(U \leq u|W = w) = \mathbb{E}[F_{U|Z}(u|Z)|W = w]$ but they do not help for identification since they introduce as many unknowns as new equations.

$$
\begin{aligned}
F_{U|Z}(u|z) &= \sum_{d=1}^{|\mathcal{D}|} \mathbb{P}(Y \leq f(d,w,u), D = d | W = w, Z = z) \\
u &= \sum_{z=1}^{|\mathcal{Z}|} F_{U|Z}(u|z)\mathbb{P}(Z = z),
\end{aligned} \tag{7}
$$

for all $u \in [0,1]$, $w \in \mathcal{W}$, and $z \in \mathcal{Z}$. For each $u \in [0,1]$, we obtain a system of $|\mathcal{Z}| \times |\mathcal{W}| + 1$ equations, for $|\mathcal{Z}| + |\mathcal{D}| \times |\mathcal{W}|$ unknowns: the $|\mathcal{D}| \times |\mathcal{W}|$ outcome functions' values $f(d,w,u)$, the $|\mathcal{Z}|$ conditional probabilities $F_{U|Z}(u|z)$. Provided that there are more equations than the number of unknowns, we can proceed to the identification of our objects of interest by solving this system of equations.

Let us provide a sufficient identification condition. For an integer N, let $\mathcal{I}_N$ be the identify matrix of size N. Define

$$
p_{d|u,w,z} = \begin{cases} \mathbb{P}(D = d | U = u, W = w, Z = z) \text{ if } f_{U|Z}(u|z) > 0, \\ 0 \text{ otherwise if } f_{U|Z}(u|z) = 0, \end{cases}
$$

the selection probabilities.[7]

Assumption 2.1 (Relevance with discrete D)**.** *The*

$$
(1 + |\mathcal{Z}| \times |\mathcal{W}|) \times (|\mathcal{Z}| + |\mathcal{D}| \times |\mathcal{W}|)
$$

matrix of selection probabilities

$$
M(u) = \begin{bmatrix}
\mathcal{P}_{W1}(u) & 0 & \dots & 0 & -\mathcal{I}_{|\mathcal{Z}|} \\
0 & \mathcal{P}_{W2}(u) & \dots & \vdots & -\mathcal{I}_{|\mathcal{Z}|} \\
\vdots & \dots & \ddots & \vdots & \vdots \\
0 & \dots & \dots & \mathcal{P}_{W|\mathcal{W}|}(u) & -\mathcal{I}_{|\mathcal{Z}|} \\
0 & \dots & \dots & 0 & \mathcal{P}_Z
\end{bmatrix},
$$

$$
\textit{where} \quad \mathcal{P}_{Ww}(u) = \begin{bmatrix}
p_{1|u,w,1} & p_{2|u,w,1} & \cdots & p_{|\mathcal{D}||u,w,1} \\
p_{1|u,w,2} & p_{2|u,w,2} & \cdots & p_{|\mathcal{D}||u,w,2} \\
\vdots & \vdots & \vdots & \vdots \\
p_{1|u,w,|\mathcal{Z}|} & p_{2|u,w,|\mathcal{Z}|} & \cdots & p_{|\mathcal{D}||u,w,|\mathcal{Z}|}
\end{bmatrix} \quad \textit{is } |\mathcal{Z}| \times |\mathcal{D}|
$$

[7] We choose this definition when $f_{U|Z}(u|z) = 0$ in order for the identification proof to hold generally, even when the support of $U|Z$ is not $[0,1]$ for all $z \in \mathcal{Z}$.

$$\text{and} \quad \mathcal{P}_Z = \Big[\mathbb{P}(Z=1) \quad \mathbb{P}(Z=2) \quad \dots \quad \mathbb{P}(Z=|\mathcal{Z}|)\Big]$$

has full column rank for all $u \in [0,1]\backslash\mathcal{K}$, where $\mathcal{K} \subset [0,1]$ is a (possibly empty) finite set containing $K \geq 0$ isolated values u_k, $k = 1, \dots, K$, at which there is a rank-one deficiency, i.e., rank$\big(M(u_k)\big) = |\mathcal{Z}| + |\mathcal{D}| \times |\mathcal{W}| - 1$.

Theorem 1 (Identification)**.** *Suppose that the support of (W, Z) is $\mathcal{W} \times \mathcal{Z}$. Then, under regularity conditions that $p_{d|u=0,w,z} > 0$ for all d, w, z, there exists a unique set of strictly increasing functions, $f(d, w, \cdot)$, mapping $[0,1]$ into the support of Y given $D = d, W = w$ (for each $d \in \mathcal{D}$ and $w \in \mathcal{W}$) and conditional distribution functions $F_{U|Z}$ solving the system of equations (7).*

A sketch of the proof of the theorem can be found in Appendix A. To show the result, we first differentiate System (7) with respect to u. This yields a quasilinear system of first order differential equations. Using the Picard-Lindelöf Theorem for system of nonlinear equations, we have that this system has a unique solution under a full rank conditon along the "optimal path" (i.e., at the true outcome functions and densities) equivalent to Assumption 2.1.[8] The relevance assumption can be expressed in terms of conditional selection probabilities because these can be directly related to the joint densities of Y, D given W, Z at the optimal choices.

Identification when $|\mathcal{D}| = |\mathcal{Z}| = |\mathcal{W}| = 2$. The full column rank identification Assumption 2.1 is a relevance condition on the effect of the variables on the true selection probabilities given the unobservable U. Let us focus on the simpler case where D and W are binary with $\mathcal{D} = \mathcal{W} = \{1, 2\}$. In this case, a direct necessary condition is that $|\mathcal{Z}| \geq 3$ for Assumption 2.1 to hold, because otherwise the system would have less equations than unknowns, so M would not be injective. Let us assume $|\mathcal{Z}| = 3$. Then, we have $|\mathcal{Z}| \times |\mathcal{W}| + 1 = 7$ equations for $|\mathcal{Z}| + |\mathcal{D}| \times |\mathcal{W}| = 7$ unknowns (four functions, and three conditional distribution functions), and one can show that

$$\begin{aligned} \det\big(M(u)\big) = {} & p_{1|u,1,1}\big(p_{1|u,2,2} - p_{1|u,2,3}\big) \\ & + p_{1|u,1,2}\big(p_{1|u,2,3} - p_{1|u,2,1}\big) + p_{1|u,1,3}\big(p_{1|u,2,1} - p_{1|u,2,2}\big). \end{aligned} \tag{8}$$

[8] Exploiting the known monotonicity of the potential outcomes functions we want to identify, we show that identification can be obtained even with some isolated violations (rank one deficiency) of the full rank condition on the optimal path, as in Bruneel-Zupanc (2023) or Feng (2024).

Hence, $\det\big(M(u)\big) = 0$, if $W \perp\!\!\!\perp D|Z,U$, or, $Z \perp\!\!\!\perp D|W,U$. In particular, it means that Z must be relevant for D but not only because of its dependence with U. It also requires that W and Z are jointly relevant, in the sense their interaction has a significant effect on the selection. For example, if $\mathbb{P}(D=1|W=w,Z=z,U=u) = a+b\mathbb{1}(w=2)+c\mathbb{1}(z=2)+d\mathbb{1}(z=3)$, there is no joint effect of W and Z on D, and one can check that the relevance condition is violated. If we add some interaction terms, e.g., $e\mathbb{1}(w=1,z=2)$, the relevance holds.

2.3.3 Identification with continuous D

Let us now consider identification with continuous D. We fix $u \in [0,1]$ and introduce $\epsilon = Y - f(D,W,u)$. Let $f_{\epsilon|D,Z,W}(e|d,z,w)$ be the density of ϵ given $D=d, Z=z, W=w$ at the point e (which existence is guaranteed under our assumptions). We also introduce $\mathcal{H} : \{h : \mathcal{Z} \to \mathbb{R} : \ \mathbb{E}[h(Z)] = 0\}$. For $\Delta \in L^2(D,W)$, we define

$$\omega_\Delta(D,Z,W) = \int_0^1 f_{\epsilon|D,Z,W}(\delta\Delta(D,W)|D,Z,W)d\delta.$$

Let us state the following strong completeness condition and the related identification result.

Assumption 2.2 (Relevance with continuous D)**.** *For all $\Delta \in L^2(D,W)$ and $h \in \mathcal{H}$,*

$$\mathbb{E}\left[\Delta(D,W)\omega_\Delta(D,Z,W)|\, Z,W\right] + h(Z) = 0 \ a.s. \Rightarrow \Delta(D,W) = h(Z) = 0 \ a.s.$$

Theorem 2. *Assume that $(d,w) \in \mathcal{D} \times \mathcal{W} \to f(d,w,u)$ belongs to $L^2(D,W)$ and that the distibution of ϵ given D,Z,W is continuous. Let also Assumption (2.2) hold. Then, $\psi \in L^2(D,W)$ and $h \in \mathcal{H}$ solve the system (5)-(6) if and only if $\psi(D,W) = q_D(W,u)$ and $h(Z) = F_{U|Z}(u|Z)$ a.s.*

Assumption 2.2 is our model's counterpart of Assumption $L1^*$ in Appendix C of Chernozhukov and Hansen (2005). As argued in the literature relevance conditions for nonseparable models with endogenous continuous treatment are difficult to interpret, see the discussions in Canay et al. (2013); Beyhum et al. (2023) for the IVQR model. Our condition suffers from the same drawback, that is it does not possess a straightforward meaning.

3 Additive model

Let us now consider an additive model with homogenous treatment effects. Such a model is interesting because identification results can be obtained in a single framework allowing for both discrete and continuous D. Moreover, the global identification results for continuous D can be derived under more interpretable conditions. We let

$$Y_d = f(d, W, X) + U, \; \mathbb{E}[U] = 0. \tag{9}$$

This is the counterpart of equation (1) in mean regression contexts. The residual U does not depend on d, meaning that the treatment effects are homogenous (this is the analogue of the rank invariance assumption in the present additive context). We then impose

$$U = g(Z, X) + V, \; \mathbb{E}[V|Z, W, X] = 0, \tag{10}$$

which is the analogue of (2). Here, for $w \in \mathcal{W}, z \in \mathcal{Z}, x \in \mathcal{X}$, $f(d, w, x) + g(z, x)$ is the average of Y_d given $Z = z, W = w, X = x$.

Let us now turn to identification. We again suppress the dependence on X for convenience. We want to identify the sum $f(d, w) + g(z)$. Let us define the infinite dimensional parameter spaces $\mathcal{P}_f$ and $\mathcal{P}_g$ for f and g, respectively. As an example, $\mathcal{P}_f$ can be the set of square integrable, bounded or linear functions of the support $\mathcal{S}(D, W)$ of (D, W) to $\mathbb{R}$. From the model, we obtain the following system of equations

$$\begin{aligned} \mathbb{E}[Y|W, Z] &= \mathbb{E}[f(D, W) + g(Z)|W, Z] \text{ a.s.}; \\ \mathbb{E}[g(Z)] &= 0. \end{aligned} \tag{11}$$

The mappings f and g are identified from the joint distribution of (Y, D, Z, W) if this system has a unique solution.[9] This will be the case under the following completeness assumption:

Assumption 3.1 (Additive completeness)**.** *For all $f \in \mathcal{P}_f$ and $g \in \mathcal{P}_g$,*

$$\mathbb{E}[f(D, W) + g(Z)|W, Z] = \mathbb{E}[g(Z)] = 0 \; a.s. \; \Rightarrow f(D, W) = g(Z) = 0 \; a.s..$$

Completeness assumptions are standard in the literature on nonparametric instrumental variable models, see Newey and Powell (2003); Darolles et al. (2011). In this

[9] As in the quantile case, there are additional equations available, of the form $\mathbb{E}[g(Z)|W] = 0$ a.s.. However, again, such equations are not useful for identification.

literature, D is the endogenous variable, Z is the IV and the completeness assumption is different since it corresponds to $\mathbb{E}[f(D)|Z] = 0$ a.s., $\Rightarrow f(D) = 0$ a.s., for all $f \in L^2(D)$. Without the separability of the model, the needed relevance would be

$$\mathbb{E}[h(D, W, Z)|W, Z] = 0 \text{ a.s. } \Rightarrow h(D, W, Z) = 0 \text{ a.s.},$$

for functions h in a defined parameter space. This can only hold under degenerate joint distribution of (D, W, Z) or degenerate parameter space. Therefore, the separability of the model allows us to rely on a weaker condition. To further interpret Assumption 3.1, we consider two examples in the following Lemmas.

Lemma 1. *Suppose D, Z and W are discrete with respective supports $\{1, \ldots, |\mathcal{D}|\}$, $\{1, \ldots, |\mathcal{Z}|\}$, $\{1, \ldots, |\mathcal{W}|\}$. For $(d, z, w) \in \mathcal{D} \times \mathcal{Z} \times \mathcal{W}$, let $p_{d|w,z} = \mathbb{P}(D = d|W = w, Z = z)$ and $p_z = \mathbb{P}(Z = z)$. Define the parameter spaces as $\mathcal{P}_f = \{f : \mathcal{D} \times \mathcal{W} \to \mathbb{R}\}$ and $\mathcal{P}_g = \{g : \mathcal{Z} \to \mathbb{R}\}$. Assume that the following $(1+|\mathcal{Z}|\times|\mathcal{W}|)\times(|\mathcal{Z}|+|\mathcal{D}|\times|\mathcal{W}|)$ matrix of selection probabilities*

$$M = \begin{bmatrix} \mathcal{P}_{W1} & 0 & \ldots & 0 & \mathcal{I}_{|\mathcal{Z}|} \\ 0 & \mathcal{P}_{W2} & \ldots & \vdots & \mathcal{I}_{|\mathcal{Z}|} \\ \vdots & \ldots & \ddots & \vdots & \vdots \\ 0 & \ldots & \ldots & \mathcal{P}_{W|\mathcal{W}|} & \mathcal{I}_{|\mathcal{Z}|} \\ 0 & \ldots & \ldots & 0 & \mathcal{P}_Z \end{bmatrix},$$

$$\textit{with} \quad \mathcal{P}_{Ww} = \begin{bmatrix} p_{1|w,1} & p_{2|w,1} & \cdots & p_{|\mathcal{D}||w,1} \\ p_{1|w,2} & p_{2|w,2} & \cdots & p_{|\mathcal{D}||w,2} \\ \vdots & \vdots & \vdots & \vdots \\ p_{1|w,|\mathcal{Z}|} & p_{2|w,|\mathcal{Z}|} & \cdots & p_{|\mathcal{D}||w,|\mathcal{Z}|} \end{bmatrix} \textit{ is } |\mathcal{Z}| \times |\mathcal{D}|,$$

$$\textit{and} \quad \mathcal{P}_Z = \begin{bmatrix} \mathbb{P}(Z = 1) & \mathbb{P}(Z = 2) & \ldots & \mathbb{P}(Z = |\mathcal{Z}|) \end{bmatrix}$$

has full column rank. Then, Assumption 3.1 holds.

Lemma 2. *Let $f_1, f_2, \ldots, f_{r_f}$ and $g_1, g_2, \ldots, g_{r_g}$ form orthonormal families of functions of $L^2(D, W)$ and $L^2(Z)$, respectively. We consider the parameter spaces:*

$$\mathcal{P}_f = \left\{ \sum_{j=1}^{r_f} \beta_j f_j, \ \beta \in \mathbb{R}^{r_f} \right\};$$

$$\mathcal{P}_g = \left\{ \sum_{j=1}^{r_g} \alpha_j g_j, \ \alpha \in \mathbb{R}^{r_g} \right\}.$$

Suppose that there exists orthonormal functions $h_1, \ldots, h_{r_h}$ in $L^2(W, Z)$ such that the $(r_h + 1) \times (r_f + r_g)$ matrix

$$M = \begin{bmatrix} e^f_{1|1} & e^f_{2|1} & \ldots & e^f_{r_f|1} & e^g_{1|1} & e^g_{2|1} & \ldots & e^g_{r_g|1} \\ e^f_{1|2} & e^f_{2|2} & \ldots & e^f_{r_f|2} & e^g_{1|2} & e^g_{2|2} & \ldots & e^g_{r_g|2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ e^f_{1|r_h} & e^f_{2|r_h} & \ldots & e^f_{r_f|r_h} & e^g_{1|r_h} & e^g_{2|r_h} & \ldots & e^g_{r_g|r_h} \\ 0 & 0 & \ldots & 0 & \mathbb{E}[g_1(Z)] & \mathbb{E}[g_2(Z)] & \ldots & \mathbb{E}[g_{r_h}(Z)] \end{bmatrix},$$

where $\quad e^f_{j|k} = \mathbb{E}[f_j(D, W) h_k(W, Z)]$ *and* $e^g_{j|k} = \mathbb{E}[g_j(Z) h_k(W, Z)]$,

has rank equal to $r_f + r_g$. Then, Assumption 3.1 holds.

The rank condition in Lemma 1 is the counterpart of Assumption 2.1 in the additive model. It can be interpreted similarly to Assumption 2.1. Lemma 2 allows D to be continuous. It considers parameter spaces consisting in flexible linear combinations of functions. By letting r_f and r_g go to infinity one can approach nonparametric parameter spaces. An important special case of Lemma 2 is the linear case, which we elaborate on below.

Linear case. Take $\mathcal{P}_f = \{(d, w) \in \mathcal{D} \times \mathcal{W} \mapsto \alpha + \beta_D d + \beta_W w\}$ and $\mathcal{P}_g = \{z \in \mathcal{Z} \mapsto \alpha_Z + \beta_Z z\}$ and $h_1(z, w) = 1$, $h_2(z, w) = z$, $h_3(z, w) = w$ and $h_4(z, w) = zw$. Then, by (11), we have

$$\mathbb{E}[\{Y - ((\alpha + \alpha_Z) + \beta_D D + \beta_W W + \beta_Z Z)\} h_j(Z, W)] = 0, \ j = 1, \ldots, 4.$$

These are exactly the moment conditions underlying the two-stage least squares estimator using a constant, D, W and Z as independent variables and a constant, Z, W and ZW as instrumental variables. Hence, as claimed in the introduction, in this linear case, we use the interaction ZW as an instrument for D.

We conclude this section by formally stating our identification theorem.

Theorem 3. *Let Assumption 3.1 hold, then $\tilde{f} \in \mathcal{P}_f$ and $\tilde{g} \in \mathcal{P}_g$ solve (11) if and only if $\tilde{f}(D, W) = f(D, W)$ and $\tilde{g}(Z) = g(Z)$ almost surely.*

4 LATE with invalid IVs

Let us now discuss how we can identify LATEs with invalid IVs. Our LATE framework is neither nested by, nor nesting the previous nonseparable model with rank invariance, and the identification arguments differ, so we discuss this in a separate section. However, as in the previous section, the intuition remains similar: relaxing the exclusion and exogeneity restrictions requires to satisfy a (testable) stronger relevance condition than standard IVs.

4.1 LATE model

Throughout this section, let us focus on the binary case, where $D = 0$ or $D = 1$. Let us denote by Z an excluded quasi-IV with support $\mathcal{Z} \subset \mathbb{R}$, and W a complementary exogenous quasi-IV with support $\mathcal{W} \subset \mathbb{R}$. The arguments also naturally extend when W and Z contain several variables, but we do not address it for simplicity of exposition. We also introduce covariables X the standard covariates. In the spirit of Imbens and Angrist (1994), define $D(w, z, x)$ the potential D as a function of $Z = z$, $W = ws$ and $X = x$. Also denote by Y_{dw} the potential outcome under treatment d and exogenous quasi-IV equal to w. We have the following model for the (continuous or discrete) potential outcomes for $d \in \{0, 1\}$,

$$Y_{dw} = f_{dw}(X) + U_d, \text{ for all } w \in \mathcal{W}, \tag{12}$$

and a selection equation with a latent index structure with additive separability of the shocks:

$$\begin{aligned} & D^*(W, Z, X) = \mu(W, Z, X) - \eta, \\ \text{and} \quad & D(W, Z, X) = 1 \text{ if } D^*(W, Z, X) \geq 0, \ D(W, Z, X) = 0 \text{ otherwise.} \end{aligned} \tag{13}$$

We observe (Y, D, W, Z, X), where $Y = Y_{DW}$ and $D = D(W, Z, X)$. As in the standard LATE framework (Imbens and Angrist, 1994; Vytlacil, 2002; Heckman and Vytlacil, 2005), (η, U_0, U_1) are general unobserved random variables which may be correlated, yielding endogenous selection. Z does not enter directly the outcome equations, but it is (possibly) entering indirectly through U_0 and U_1. Otherwise, we impose the same assumptions (naturally adjusted for inclusion of W and endogeneity of Z) as Heckman and Vytlacil (2005), and refer to this paper for more detailed discussions:

Assumption 4.1. *The following holds:*

(i) (Independence): (U_0, U_1, η) are independent of W conditional on Z and X.

(ii) (Rank condition): $\mu(W, Z, X)$ is a nondegenerate random variable conditional on Z and X.

(iii) The distribution of η is absolutely continuous w.r.t. Lebesgue measure.

(iv) (Finite means) The values of $\mathbb{E}|Y_0|$ and $\mathbb{E}|Y_1|$ are finite.

(v) $0 < Pr(D = 1|X) < 1$.

Under independence (i), the separability of η from $\mu(W, Z, X)$ in the latent index selection equation (13) yields the monotonicity assumption of Imbens and Angrist (1994) for the exogenous quasi-IV, W. In other words, conditional on $X = x$ and $Z = z$, either $D(w', z, x) \geq D(w, z, x$ for all $(w, w') \in \mathcal{W} \times \mathcal{W}$ or $D(w', z, x) \leq D(w, z, x)$ for all $(w, w') \in \mathcal{W} \times \mathcal{W}$. Crucially, without further assumption, the monotonicity only holds for the exogenous quasi-IV, not for the endogenous one. Indeed, Z is generally correlated with η, so we do not have additive separability of the effect of Z from the shock η on D^*. As a consequence, we only express the LATE for changes of W, not changes of Z, because without monotonicity with respect to Z, the set of compliers is not easily defined. A-(ii) is a general relevance condition but we impose stronger and more precise restrictions later.

A few remarks are in order. Note that if we assume that $f_{dw}(\cdot)$ does not depend on w, then W becomes a valid IV. Similarly, Z may be independent of the shocks (U_0, U_1, η) and, in that case it is a valid IV. Remark also that here U_d does not depend on w. Hence, we assume that the effect of w on the potential outcomes is homogenous. In other words, $Y_{dw'} - Y_{dw}$ is not random. Our procedure is therefore robust to homogenous direct effects of W on the potential outcomes. Finally, we stress again that we do not assume that Z is exogenous here.

Define the propensity score $P(W, Z, X) = \mathbb{P}(D = 1|W, Z, X)$. Under our assumptions, we have that $P(W, Z, X) = F_{\eta|W,Z,X}(\mu(W, Z, X)) = F_{\eta|Z,X}(\mu(W, Z, X))$ by independence, and similarly, $P(W, Z, X) = \mathbb{P}(D(W, Z, X) = 1|W, Z, X) = \mathbb{P}(D(W, Z, X) = 1|Z, X)$ by independence. Let us normalize $\eta|Z, X \sim \mathcal{U}(0, 1)$, then, $\mu(W, Z, X) =$

$P(W, Z, X)$.[10] The setup implies that, conditional on $Z = z$ and $X = x$, for any $w' \neq w$, if $P(w', z, x) = P(w, z, x)$, then $D(w', z, x) = D(w, z, x)$ almost surely. This is the key property implied by monotonicity that we will use for identification.

4.2 Definition of the LATE with invalid IVs

From here onwards, let us abstract from X to simplify the exposition, and proceed as if every statement was conditional on X. For any $z \in \mathcal{Z}$, conditional on $Z = z$, for any w' and $w \in \mathcal{W}$, with $P(w, z) = p$ and $P(w', z) = p'$, define the counterpart of Imbens and Angrist (1994)'s LATE with invalid IVs as

$$\Delta_{LATE}(w, w'|z) = \mathbb{E}[\ Y_{1w'} - Y_{0w} \mid D(w', z) = 1, D(w, z) = 0, Z = z\ ]. \tag{14}$$

In other words, this LATE is the average variation in outcome following from an exogenous change in W from w to w', when $Z = z$, for subjects switching from no treatment to treatment because of this change. As already explained, we can only define a LATE conditional on Z, since monotonicity does not hold with respect to Z, it is not possible to identify the effect exogenously moving Z on the subset of the population taking up treatment because of this change. In a sense, Z behaves similarly to a standard covariates, X, in the definition of the LATE. However, its (conditional) exclusion from Y is going to allow us to identify the said LATE. Finally, remark also that if W is in fact excluded, (14) becomes the original LATE of Imbens and Angrist (1994) (conditional on Z).

4.3 Identification of the LATE

Let us study identification $\Delta_{LATE}(w, w'|z)$ using data on (D, Y, W, Z). For any w and $w' \in \mathcal{W}$ and $z \in \mathcal{Z}$, with $P(w, z) = p < P(w', z) = p'$, we can compute

$$\begin{aligned}
&\mathbb{E}\big[Y|W = w', Z = z\big] - \mathbb{E}\big[Y|W = w, Z = z\big] \\
&= \mathbb{E}\big[Y_{0w'} + (Y_{1w'} - Y_{0w'})D|W = w', Z = z\big] - \mathbb{E}\big[Y_{0w} + (Y_{1w} - Y_{0w})D|W = w, Z = z\big] \\
&= \mathbb{E}\big[Y_{0w'} + (Y_{1w'} - Y_{0w'})D(w', z)|Z = z\big] - \mathbb{E}\big[Y_{0w} + (Y_{1w} - Y_{0w})D(w, z)|Z = z\big]
\end{aligned}$$

[10]This normalization is innocuous given the model assumptions, if the latent variable generating the choices is, $D^* = \nu(W, Z, X) - V$, where V is a general continuous random variable, we can always reparametrize the model such that $\mu(W, Z, X) = F_{V|Z,X}(\nu(W, Z, X))$ and $\eta = F_{V|Z,X}(V)$.

$$
\begin{aligned}
&= \mathbb{E}\big[(Y_{1w'} - Y_{0w})(D(w',z) - D(w,z))|Z=z\big] \quad \Big\} = \textit{ effect on compliers} \\
&\quad + \underbrace{\mathbb{E}\big[(Y_{0w'} - Y_{0w})(1 - D(w',z))|Z=z\big]}_{\textit{effect on never takers}} \\
&\quad + \underbrace{\mathbb{E}\big[(Y_{1w'} - Y_{1w})D(w,z)|Z=z\big]}_{\textit{effect on always takers}}, \qquad (15)
\end{aligned}
$$

where the second equality comes from the independence of W conditional on Z (A-1), and the third one comes from rewriting $Y_{1w'} - Y_{0w'} = Y_{1w'} - Y_{0w} + Y_{0w} - Y_{0w'}$, and $Y_{1w} - Y_{0w} = Y_{1w} - Y_{1w'} + Y_{1w'} - Y_{0w}$. Moreover, by monotonicity/uniformity, $D(w',z) > D(w,z)$ for all subjects if $p' > p$. Thus, the effect on compliers is almost directly the LATE, since:

$$
\begin{aligned}
&\mathbb{E}\big[(Y_{1w'} - Y_{0w})(D(w',z) - D(w,z))|Z=z\big] \\
&= \mathbb{E}\big[Y_{1w'} - Y_{0w}|D(w',z) = 1, D(w,z) = 0, Z = z]\Pr(D(w',z) = 1, D(w,z) = 0) \\
&= \Delta_{LATE}(w,w'|z)(p' - p).
\end{aligned}
$$

The main difference with the standard LATE with valid IVs is the presence of the last two terms in Equation (15). They represent the effects on always takers and on never takers following from an exogenous change in W from w to w' conditional on $Z = z$. These terms are present because W is (possibly) included here, so changing W does not only affect the compliers, but also the never-takers and always-takers.

To identify the LATE with invalid IVs, one needs to recover these two additional effects from the data. Let us focus on the effect on always takers for example. We want to identify

$$
\begin{aligned}
\mathbb{E}\big[(Y_{1w'} - Y_{1w})D(w,z)|Z=z\big] &= (f_{1w'} - f_{1w})\ \mathbb{E}[D(w,z)|Z=z] \\
&:= \Delta_1^W(w,w')\ p.
\end{aligned}
$$

So, if we identify $\Delta_1^W(w,w')$, we identify the effect on the always takers. Similarly, define $\Delta_0^W(w,w') = f_{0w'} - f_{0w}$. If we identify $\Delta_0^W(w,w')$, we identify the effect on the never-takers.

We can show that $\Delta_d^W(w,w')$ (for $d = 0,1$) are identified under the following irrelevance assumption.

Assumption 4.2 (Local Irrelevance)**.** *There exists $Z = z^*$ such that, $P(w,z^*) =$*

$P(w', z^*)$.

The local irrelevance assumption requires that there exists z^* such that the effect effect of W and Z on D is sufficiently nonlinear. Such an assumption is more likely to hold when Z is continuous, but it does not require that W or Z are continuous. As mentioned in the introduction, in a different context, D'Haultfœuille et al. (2021) also relies on a local irrelevance condition, and show the power of such restriction. We also note that Assumption 4.2 is testable.

Under Local Irrelevance, 4.2, we identify $\Delta_1^W(w, w')$ and $\Delta_0^W(w, w')$, and consequently, we identify $\Delta_{LATE}(w, w', z)$ for all $z \in \mathcal{Z}$. Indeed, at $Z = z^*$ such that $P(w, z^*) = P(w', z^*) = p^*$, we have $D(w, z^*) = D(w', z^*)$ conditional on $Z = z^*$ by monotonicity. As a consequence, we have the following theorem.

Theorem 4. *Let Assumptions 4.1 and 4.2 hold. Suppose that* $0 < P(w, z^*) < 1$, *where* z^* *is defined in Assumption 4.2. Then,* $\Delta_{LATE}(w, w'|z)$*is identified for all* $z \in \mathcal{Z}$ *such that* $P(w', z) - P(w, z) > 0$ *and* $\Delta_{LATE}(w', w|z)$*is identified for all* $z \in \mathcal{Z}$ *such that* $P(w', z) - P(w, z) < 0$.

Special case of fuzzy DiD. In the case where $W \in \{0, 1\}$ is time (with two time periods) and $Z \in \{0, 1\}$ is the group assignment (with two groups), it is helpful to compare our results with that of De Chaisemartin and d'Haultfoeuille (2018) for fuzzy DiD. They make two crucial sets of assumptions. First, there are the common trends and conditional common trends conditions (Assumptions 2 and 4' in their paper). These assumptions are both implied by our model (12). Formally, our model (12) is more restrictive, but in the special case of fuzzy DiD, we could relax it to just impose some conditions on the trends as De Chaisemartin and d'Haultfoeuille (2018). Second, they assume that the treatment rate does not change in the control group between the two periods (Assumption 2 in their paper). This is just our local irrelevance Assumption 4.2 where $Z = z^*$ is the control group and $w = 0$, $w' = 1$ are the dates.

5 Conclusion

In this paper, we have shown that quantities typically identified by IV can be identified by two quasi-IVs instead. Several economic examples of quasi-IVs satisfying our assumptions have been given, demonstrating that our approach is promising in practice. In the simple linear case, the estimation of treatment effects with quasi-IVs can be carried by running a standard 2SLS and taking the interaction between the complementary quasi-IVs as a valid IV. However, for the more general models presented in the paper, a natural and crucial next step before bringing the methodology to the data is estimation. Our identification proofs are constructive and in this sense can be used to define an estimator. Nevertheless, there will be several computational challenges arising from the complex nature of the problem at hand that we will tackle in the future. Another avenue is to study identification with other types of invalid IVs which complement each other in a different manner from the exogenous and excluded quasi-IVs studied here. The semi-IVs of Bruneel-Zupanc (2023) are one example, but they may be other empirically relevant cases.

References

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings," *Econometrica*, 70, 91–117.

Ackerberg, D. A., K. Caves, and G. Frazer (2015): "Identification properties of recent production function estimators," *Econometrica*, 83, 2411–2451.

Adao, R., M. Kolesár, and E. Morales (2019): "Shift-share designs: Theory and inference," *The Quarterly Journal of Economics*, 134, 1949–2010.

Ambrosetti, A. and G. Prodi (1995): *A primer of nonlinear analysis*, 34, Cambridge University Press.

Angrist, J., E. Bettinger, E. Bloom, E. King, and M. Kremer (2002): "Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment," *American economic review*, 92, 1535–1558.

Angrist, J. D. (1990): "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *The American Economic Review*, 80, 313–336.

Athey, S. and G. W. Imbens (2006): "Identification and inference in nonlinear difference-in-differences models," *Econometrica*, 74, 431–497.

Ban, K. and D. Kédagni (2022): "Nonparametric bounds on treatment effects with imperfect instruments," *The Econometrics Journal*, 25, 477–493.

Bartels, L. M. (1991): "Instrumental and "Quasi-Instrumental" Variables," *American Journal of Political Science*, 35, 777–800.

Bartik, T. J. (1991): "Who benefits from state and local economic development policies?" *WE Upjohn Institute for Employment Research.*

Beyhum, J., S. Centorrino, J.-P. Florens, and I. Van Keilegom (2023): "Instrumental variable estimation of dynamic treatment effects on a duration outcome," *Journal of Business & Economic Statistics*, 1–11.

Bloom, H. S., L. L. Orr, S. H. Bell, G. Cave, F. Doolittle, W. Lin, and J. M. Bos (1997): "The benefits and costs of JTPA Title II-A programs: Key findings from the National Job Training Partnership Act study," *Journal of human resources*, 549–576.

Blundell, R. and S. Bond (2000): "GMM Estimation with persistent panel data: an application to production functions," *Econometric Reviews*, 19, 321–340.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016a): "Female Labor Supply, Human Capital, and Welfare Reform," *Econometrica*, 84, 1705–1753.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2016b): "Consumption Inequality and Family Labor Supply," *The American Economic Review*, 106, 387–435.

Borusyak, K. and P. Hull (2023): "Nonrandom Exposure to Exogenous Shocks," *Econometrica*, 91, 2155–2185.

Borusyak, K., P. Hull, and X. Jaravel (2022): "Quasi-experimental shift-share research designs," *The Review of Economic Studies*, 89, 181–213.

Bruneel-Zupanc, C. (2022): "Discrete-Continuous Dynamic Choice Models: Identification and Conditional Choice Probability Estimation," *SSRN Electronic Journal*.

——— (2023): "Don't (fully) exclude me, it's not necessary! Identification with semi-IVs," *arXiv preprint arXiv:2303.12667*.

Caetano, C. and J. C. Escanciano (2021): "Identifying multiple marginal effects with a single instrument," *Econometric Theory*, 37, 464–494.

Canay, I. A., A. Santos, and A. M. Shaikh (2013): "On the testability of identification in some nonparametric models with endogeneity," *Econometrica*, 81, 2535–2559.

Card, D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*.

——— (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69, 1127–1160.

Card, D. and A. B. Krueger (1994): "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review*, 84, 772–793.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011): "Estimating Marginal Returns to Education," *The American Economic Review*, 101, 2754–2781.

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

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

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011): "Nonparametric instrumental regression," *Econometrica*, 79, 1541–1565.

De Chaisemartin, C. and X. d'Haultfoeuille (2018): "Fuzzy differences-in-differences," *The Review of Economic Studies*, 85, 999–1028.

De Loecker, J. and C. Syverson (2021): "Chapter 3 - An industrial organization perspective on productivity," in *Handbook of Industrial Organization, Volume 4*, ed. by K. Ho, A. Hortaçsu, and A. Lizzeri, Elsevier, vol. 4 of *Handbook of Industrial Organization*, 141–223.

Dong, Y. (2010): "Endogenous regressor binary choice models without instruments, with an application to migration," *Economics Letters*, 107, 33–35.

D'Haultfœuille, X., S. Hoderlein, and Y. Sasaki (2021): "Testing and relaxing the exclusion restriction in the control function approach," *Journal of Econometrics*.

Escanciano, J. C., D. Jacho-Chávez, and A. Lewbel (2016): "Identification and estimation of semiparametric two-step models," *Quantitative Economics*, 7, 561–589.

Feng, J. (2024): "Matching points: Supplementing instruments with covariates in triangular models," *Journal of Econometrics*, 238, 105579.

Flores, C. A. and A. Flores-Lagunes (2013): "Partial identification of local average treatment effects with an invalid instrument," *Journal of Business & Economic Statistics*, 31, 534–545.

Gandhi, A., S. Navarro, and D. A. Rivers (2020): "On the identification of gross output production functions," *Journal of Political Economy*, 128, 2973–3016.

Gao, W. Y. and R. Wang (2023): "IV Regressions without Exclusion Restrictions," .

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020): "Bartik instruments: What, when, why, and how," *American Economic Review*, 110, 2586–2624.

Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017–1098.

Heckman, J. J., J. Smith, and N. Clements (1997): "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts," *The Review of Economic Studies*, 64, 487–535.

Heckman, J. J. and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation1," *Econometrica*, 73, 669–738.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Jiang, F. and E. S. Tsyawo (2022): "A Consistent ICM-based χ^2 Specification Test," *arXiv preprint arXiv:2208.13370*.

Klein, R. and F. Vella (2010): "Estimating a class of triangular simultaneous equations models without exclusion restrictions," *Journal of Econometrics*, 154, 154–164.

Levinsohn, J. and A. Petrin (2003): "Estimating production functions using inputs to control for unobservables," *The review of economic studies*, 70, 317–341.

LEWBEL, A. (2012): "Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models," *Journal of Business & Economic Statistics*, 30, 67–80.

——— (2018): "Identification and estimation using heteroscedasticity without instruments: The binary endogenous regressor case," *Economics Letters*, 165, 10–12.

LEWBEL, A., S. M. SCHENNACH, AND L. ZHANG (2023): "Identification of a triangular two equation system without instruments," *Journal of Business & Economic Statistics*, 1–12.

LIU, S., I. MOURIFIÉ, AND Y. WAN (2020): "Two-way exclusion restrictions in models with heterogeneous treatment effects," *The Econometrics Journal*, 23, 345–362.

MANSKI, C. F. AND J. V. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010.

MARSZALEK, W., T. AMDEBERHAN, AND R. RIAZA (2005): "Singularity crossing phenomena in DAEs: a two-phase fluid flow application case study," *Computers and Mathematics with Applications*, 49, 303–319.

MEALLI, F. AND B. PACINI (2013): "Using secondary outcomes to sharpen inference in randomized experiments with noncompliance," *Journal of the American Statistical Association*, 108, 1120–1131.

MROZ, T. A. (1987): "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica*, 55, 765–799.

NEVO, A. AND A. M. ROSEN (2012): "Identification with imperfect instruments," *Review of Economics and Statistics*, 94, 659–671.

NEWEY, W. K. AND J. L. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

OLLEY, G. S. AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

RIGOBON, R. (2003): "Identification through heteroskedasticity," *Review of Economics and Statistics*, 85, 777–792.

SCHOCHET, P. Z., J. BURGHARDT, AND S. MCCONNELL (2008): "Does job corps work? Impact findings from the national job corps study," *American economic review*, 98, 1864–1886.

TSYAWO, E. S. (2023): "Feasible IV regression without excluded instruments," *The Econometrics Journal*, 26, 235–256.

VYTLACIL, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341.

WANG, R. (2023): "Point Identification of LATE with Two Imperfect Instruments," .

A Sketch of the proof of Theorem 1

Identification of the (strictly increasing) outcome functions and distribution of $U|Z$ for all $u \in [0,1]$ requires that there exists a unique solution to the system (7). To find a sufficient condition for uniqueness of the solution, one can derive each equation of System (7) with respect to u and obtain a quasilinear system of first order differential equations,

$$\begin{cases} 0 & = \sum_{d=1}^{|\mathcal{D}|} f_{Y,D|W,Z}(f(d,w,u), d|w,z) \quad \partial f(d,w,u)/\partial u - f_{U|Z}(u|z) \\ 1 & = \sum_{z=1}^{|\mathcal{Z}|} f_{U|Z}(u|z)\mathbb{P}(Z=z), \end{cases} \tag{16}$$

where the unknowns are the derivatives (with respect to u) of the outcome functions, $\partial f(d,w,u)/\partial u$, and the derivatives of the conditional distributions of $U|Z$, i.e., the conditional densities $f_{U|Z}(u|z)$ for all $(w,z) \in \mathcal{W} \times \mathcal{Z}$. To shorten the notation, and minimize the risk of confusion between the outcomes and densities, let us denote $q_{dw}(u) = f(d,w,u)$, and its derivative $q'_{dw}(u) = \partial f(d,w,u)/\partial u$. Let us denote the vector of all the functions of interest,

$$\mathbf{q}(u) = \Big[q_{11}(u)\ q_{21}(u) \cdots q_{|\mathcal{D}|1}(u) \cdots q_{1|\mathcal{W}|}(u) \cdots q_{|\mathcal{D}||\mathcal{W}|}(u) \quad F_{U|Z}(u|1) \cdots F_{U|Z}(u||\mathcal{Z}|)\Big],$$

and its derivative:

$$\mathbf{q}'(u) = \Big[q'_{11}(u)\ q'_{21}(u) \cdots q'_{|\mathcal{D}|1}(u) \cdots q'_{1|\mathcal{W}|}(u) \cdots q'_{|\mathcal{D}||\mathcal{W}|}(u) \quad f_{U|Z}(u|1) \cdots f_{U|Z}(u||\mathcal{Z}|)\Big].$$

Also define the conditional joint densities as $f_{d|w,z}(y) = f_{Y,D|W,Z}(y,d|w,z)$ if $f_{U|Z}(u|z) > 0$ and $f_{d|w,z}(y) = 0$ if $f_{U|Z}(u|z) = 0$. The system of differential equation (16) can be written under matrix form as,[11]

$$\tilde{M}\Big(\mathbf{q}(u)\Big)\mathbf{q}'(u) = \Big[\underbrace{0 \cdots 0}_{\text{size } |\mathcal{D}|\times|\mathcal{W}|} \quad \underbrace{1 \cdots 1}_{\text{size } |\mathcal{Z}|}\Big]^T, \tag{17}$$

[11] Notice that the conditional distributions $F_{U|Z}(u)$ which are included in $\mathbf{q}(u)$, play no role in M. We include them nonetheless for generality of the notation.

where

$$\tilde{M}(\mathbf{q}(u)) = \begin{bmatrix} \mathcal{F}_{W1}(\mathbf{q}(u)) & 0 & \dots & 0 & -\mathcal{I}_{|\mathcal{Z}|} \\ 0 & \mathcal{F}_{W2}(\mathbf{q}(u)) & \dots & \vdots & -\mathcal{I}_{|\mathcal{Z}|} \\ \vdots & \dots & \ddots & \vdots & \vdots \\ 0 & \dots & \dots & \mathcal{F}_{W|\mathcal{W}|}(\mathbf{q}(u)) & -\mathcal{I}_{|\mathcal{Z}|} \\ 0 & \dots & \dots & 0 & \mathcal{P}_Z \end{bmatrix},$$

$$\text{with} \quad \mathcal{F}_{Ww}(\mathbf{q}(u)) = \begin{bmatrix} f_{1|w,1}(q_{1w}(u)) & f_{2|w,1}(q_{2w}(u)) & \dots & f_{|\mathcal{D}||w,1}(q_{1w}(u)) \\ f_{1|w,2}(q_{1w}(u)) & f_{2|w,2}(q_{2w}(u)) & \dots & f_{|\mathcal{D}||w,2}(q_{1w}(u)) \\ \vdots & \vdots & \vdots & \vdots \\ f_{1|w,|\mathcal{Z}|}(q_{1w}(u)) & f_{2|w,|\mathcal{Z}|}(q_{2w}(u)) & \dots & f_{|\mathcal{D}||w,|\mathcal{Z}|}(q_{1w}(u)) \end{bmatrix} \text{ is } |\mathcal{Z}| \times |\mathcal{D}|,$$

$$\text{and} \quad \mathcal{P}_Z = \Big[\mathbb{P}(Z=1) \quad \mathbb{P}(Z=2) \quad \dots \quad \mathbb{P}(Z=|\mathcal{Z}|)\Big]$$

Now, if the matrix of conditional densities, $\tilde{M}(\mathbf{q}(u))$, is full column rank, it has a left inverse, that we denote $\tilde{M}^{-1}(\mathbf{q}(u))$, and we can recover $\mathbf{q}'(u)$ as

$$\mathbf{q}'(u) = \tilde{M}^{-1}\Big(\mathbf{q}(u)\Big)\Big[0\cdots 0\ 1\cdots 1\Big]^T.$$

Starting from the known minimum $\mathbf{q}(0) = [\cdots q_{dw}(0)\cdots F_{U|Z}(0|z)\cdots]$, with $F_{U|Z}(0|z) = 0$ and $q_{dw}(0)$ identified as the minimum observable in the data (if $p_{d|u=0,w,z} > 0$ for all d, w, z), we can solve for the unique optimal solution path if $\tilde{M}(\mathbf{q}(u))$ is full column rank for all u. So a key condition for identification is the full column rank of $\tilde{M}(\mathbf{q}(u))$, which allows us to use Picard-Lindelöf Theorem for nonlinear system of differential equations.

For the interpretation, notice that the joint conditional densities at the true values of the outcome functions are related to the conditional selection probabilities as

$$f_{Y,D|W,Z}(q_{dw}(u), d|w, z) = p_{d|u,w,z} \times \frac{\partial u_{dw}}{\partial y_{dw}}(q_{dw}(u)),$$

where $u_{dw}(y_{dw}) = q_{dw}^{-1}(y_{dw})$ for all d, w. As a consequence, we can rewrite $\tilde{M}(\mathbf{q}(u))$ as

$$\tilde{M}\big(\mathbf{q}(u)\big) = M(u)\ H\big(\mathbf{q}(u)\big),$$

with the $M(u)$ defined in the relevance Assumption 2.1 and where

$$H(\mathbf{q}(u)) = \operatorname{diag}\left(\frac{\partial u_{11}}{\partial y_{11}}(q_{11}(u))\cdots\frac{\partial u_{|\mathcal{D}|1}}{\partial y_{|\mathcal{D}|1}}(q_{|\mathcal{D}|1}(u))\cdots\frac{\partial u_{|\mathcal{D}||\mathcal{W}|}}{\partial y_{|\mathcal{D}||\mathcal{W}|}}(q_{|\mathcal{D}||\mathcal{W}|}(u)) \quad \underbrace{1\cdots 1}_{\text{size } |\mathcal{Z}|}\right).$$

By strict monotonicity of the outcome functions, all the diagonal elements of $H(\mathbf{q}(u))$ are strictly positive. As a consequence, $\tilde{M}(\mathbf{q}(u))$ is full column rank if and only if $M(u)$ is full column rank. Hence, the relevance Assumption 2.1 is expressed directly in terms of the true selection probabilities. Notice that the condition only needs to hold on the true optimal path with the true outcome function and density of $U|Z$. This is because, if we start on the correct path, we will not deviate from it if the relevance condition is satisfied along the path.

The main idea behind the proof has been expressed above. Now, especially with a large number of alternative in D, it is likely even with relatively simple models, that the matrix of selection probability may have a rank one deficiency at some points along the way as it naturally occurs that some of the alternatives become uninformative at some isolated values of u. However, this is not a problem and following the proof in Appendix A of Bruneel-Zupanc (2023) building on the literature on quasilinear differential equations (Marszalek et al., 2005), we show that, thanks to the known strict monotonicity of the potential outcomes, the unique solution path is identified, even if there are an isolated set of values of u, denoted u_k at which the matrix $M(u_k)$ has a rank-one deficiency. A similar proof, building on variants of Hadamard's Theorem for global inverse (Ambrosetti and Prodi, 1995) can be found in Feng (2024). ■

B Proof of Theorem 2

Let $(\varphi, G_{Z|U})$ be another admissible solution to the system (5)-(6). The following holds

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[\mathbb{1}(Y \leq f(D,W,u)) - \mathbb{1}(Y_D \leq \varphi(D,W,u))|D,W,Z]|W,Z] \\
&\quad + F_{Z|U}(u|Z) - G_{Z|U}(u|Z) = 0 \text{ a.s.}. \\
&\mathbb{E}[F_{Z|U}(u|Z) - G_{Z|U}(u|Z)] = 0 \text{ a.s.},
\end{aligned} \tag{18}$$

Let $\Delta(D,W) = \varphi(D,W,u) - f(D,W,u)$ and $h(Z) = F_{Z|U}(u|Z) - G_{Z|U}(u|Z)$. By definition of ϵ, the first equation of (18) is equivalent to

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[I(\epsilon \leq 0) - \mathbb{E}[I(\epsilon \leq \Delta(D,W))|D,W,Z]|Z,W] + h(Z) = 0 \text{ a.s.} \\
\iff\ &\mathbb{E}\left[\int_0^1 \Delta(D,W) f_{\epsilon|D,Z,W}(\delta\Delta(D,W)|D,Z,W)\middle| W,Z\right] + h(Z) = 0 \text{ a.s.}
\end{aligned}$$

The system (18) is then equivalent to

$$\begin{aligned} \mathbb{E}\left[\Delta(D,W)\omega_\Delta(D,Z,W)|\, W,Z\right] + h(Z) = 0; \\ \mathbb{E}[h(Z)] = 0, \end{aligned} \tag{19}$$

which implies $\Delta(D,W) = g(Z) = 0$ a.s. by Assumption 2.2.

C Proof of the results of Section 3

C.1 Proof of Lemma 1

Take $\tilde{f} \in \mathcal{P}_f$ and $\tilde{g} \in \mathcal{P}_g$ such that

$$\mathbb{E}[\tilde{f}(D,W) + \tilde{g}(Z)|W,Z] = \mathbb{E}[\tilde{g}(Z)] = 0 \text{ a.s.} \tag{20}$$

Let $v = [\tilde{f}(1,1),\dots,\tilde{f}(|\mathcal{D}|,1),\dots,\tilde{f}(1,|\mathcal{W}|),\dots,\tilde{f}(|\mathcal{D}|,|\mathcal{W}|),\tilde{g}(1),\dots,\tilde{g}(|\mathcal{Z}|)]^\top$. The system of equations (20) can be rewritten as $Mv = 0$, where The full column rank assumption on M implies $v = 0$.

C.2 Proof of Lemma 2

Take $\tilde{f} = \sum_{j=1}^{r_f} \beta_j f_j \in \mathcal{P}_f$ and $\tilde{g} = \sum_{j=1}^{r_j} \beta_j g_j \in \mathcal{P}_g$ such that

$$\mathbb{E}[\tilde{f}(D,W) + \tilde{g}(Z)|W,Z] = \mathbb{E}[\tilde{g}(Z)] = 0 \text{ a.s.}$$

By the law of total expectations, this implies

$$\mathbb{E}[(\tilde{f}(D,W) + \tilde{g}(Z))h_j(W,Z)] = \mathbb{E}[\tilde{g}(Z)] = 0 \text{ a.s.},\ j = 1,\dots,r_h. \tag{21}$$

Let $v = [\beta_1,\dots,\beta_{r_f},\alpha_1,\dots,\alpha_{r_g}]^\top$. The system of equations (21) can be rewritten as $Mv = 0$. This yields $v = 0$ by the full column rank condition on M.

C.3 Proof of Theorem 3

Suppose that $\tilde{f} \in \mathcal{P}_f$ and $\tilde{g} \in \mathcal{P}_g$ satisfy

$$\begin{aligned} \mathbb{E}[Y|W,Z] &= \mathbb{E}[\tilde{f}(D,W) + \tilde{g}(Z)|W,Z] \text{ a.s.} \\ \mathbb{E}[\tilde{g}(Z)] &= 0. \end{aligned}$$

Then, we obtain

$$\mathbb{E}[\Delta f(D,W)|W,Z] + \Delta g(Z) = \mathbb{E}[\Delta g(Z)] = 0 \text{ a.s.},$$

where $\Delta f = \tilde{f} - f$ and $\Delta g = \tilde{g} - g$. This yields the result by Assumption 3.1.