WORKING PAPER · NO. 2022-142

Instrumental Variables with Unordered Treatments: Theory and Evidence from Returns to Fields of Study

Eskil Heinesen, Christian Hvid, Lars Kirkebøen, Edwin Leuven, and Magne Mogstad

OCTOBER 2022

Becker Friedman Institute | Ronzetti Initiative for the Study of Labor Markets

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

Instrumental variables with unordered treatments: Theory and evidence from returns to fields of study*

Eskil Heinesen† Christian Hvid‡ Lars Kirkebøen§
Edwin Leuven¶ Magne Mogstad‖

Abstract

We revisit the identification argument of Kirkeboen et al. (2016) who showed how one may combine instruments for multiple unordered treatments with information about individuals' ranking of these treatments to achieve identification while allowing for both observed and unobserved heterogeneity in treatment effects. We show that the key assumptions underlying their identification argument have testable implications. We also provide a new characterization of the bias that may arise if these assumptions are violated. Taken together, these results allow researchers not only to test the underlying assumptions, but also to argue whether the bias from violation of these assumptions are likely to be economically meaningful. Guided and motivated by these results, we estimate and compare the earnings payoffs to post-secondary fields of study in Norway and Denmark. In each country, we apply the identification argument of Kirkeboen et al. (2016) to data on individuals' ranking of fields of study and field-specific instruments from discontinuities in the admission systems. We empirically examine whether and why the payoffs to fields of study differ across the two countries. We find strong cross-country correlation in the payoffs to fields of study, especially after removing fields with violations of the assumptions underlying the identification argument.

*We thank Arnstein Vestre for excellent research assistance. Edwin Leuven recognizes support from the Norwegian Research Council, project no. 275906.

†Rockwool Foundation Research Unit. Email: esh@rff.dk

‡Email: christian@hvids.eu

§Statistics Norway. Email: kir@ssb.no

¶Department of Economics, University of Oslo; Statistics Norway; CESifo; IZA. Email: edwin.leuven@econ.uio.no.

‖University of Chicago, Department of Economics; Statistics Norway; NBER. Email: magne.mogstad@gmail.com

1 Introduction

Instrumental variables (IV) estimation of treatment effects is challenging if there are multiple unordered treatments. Not only does identification require (at least) one instrument per alternative, but it is also necessary to deal with the issue that individuals who choose the same treatment may have different next-best treatments. One way to resolve this issue is to assume homogeneous treatment effects. If effects are heterogeneous across individuals (conditional on observable characteristics), then standard 2SLS does not identify the payoff to any individual or group of the population from choosing one treatment instead of another.[1]

We revisit the identification argument of Kirkeboen et al. (2016) who showed how one may combine instruments for multiple unordered treatments with information about individuals' ranking of these treatment to achieve identification while allowing for both observed and unobserved heterogeneity in treatment effects.[2] We show that the key assumptions underlying their identification argument have testable implications. We also provide a new characterization of the bias that may arise if these assumptions are violated.[3] Taken together, these results allow researchers not only to test the underlying assumptions, but also to argue whether the bias from violation of these assumptions are likely to be economically meaningful. Guided and motivated by these results, we estimate and compare the earnings payoffs to post-secondary fields of study in Norway and Denmark.[4] In each country, we apply the identification argument of Kirkeboen et al. (2016) to data on individuals' ranking of fields of study and field-specific instruments from discontinuities in the admission systems. We then empirically examine the extent to which and why the payoffs to fields of study differ across the two countries.

In Section 2, we begin by briefly reviewing IV in settings with multiple unordered treatments, laying the groundwork for our analysis. As in the analysis of binary treatments in Imbens and Angrist (1994), we allow for heterogeneous effects and assume that each instrument is exogenous and satisfies a monotonicity condition. Our point of departure is

[1] A number of studies in diverse fields report evidence of unobserved heterogeneity in causal effects (see, for example, the review article by Mogstad and Torgovitsky (2018)).

[2] Kirkeboen et al. (2016) contributes to a larger literature on identification of treatment effects in unordered choice models. Heckman et al. (2006) and Heckman and Urzúa (2010) discuss the challenges associated with the identification and interpretation of treatment effects in such models. See also the recent work by Kamat (2017) and Lee and Salanié (2020).

[3] Throughout the paper, we use the term *bias* to describe the difference between two population quantities, namely the IV estimand and the parameter of interest, that is the positively weighted average of treatment effects for some complier group.

[4] There is a growing body of work on the payoffs to field of study or college major, reviewed in Altonji et al. (2012, 2016), and Kirkeboen et al. (2016). The latter study also reports IV estimates of the payoffs to fields of study from Norway. Thus, our empirical contribution is the new payoff estimates from Denmark, the examination of the IV assumptions, and the comparison of payoffs to fields of study between Norway and Denmark.

the key result in Kirkeboen et al. (2016): IV can then be used to identify local average treatment effects (LATEs) of unordered treatments under the additional assumptions that the analyst observes individuals' next-best alternatives and an irrelevance condition on preferences.

The next two sections of the paper examine whether the additional assumptions of Kirkeboen et al. (2016) have testable implications and the bias that may arise if they are violated. To do so, it is useful to stratify the population into a set of instrument-dependent groups, sometimes referred to as principal strata. These groups are defined by the manner in which members of the population react to the instruments. In addition to the usual compliers, always takers, and never takers of Imbens and Angrist (1994), there are two so-called defier groups (both of which are distinct from the usual defier group that exists if the usual monotonicity condition fails). The first is the next-best defiers. In the context of our empirical application, this group consists of individuals who would choose their preferred field if above the admission cutoff, but otherwise choose fields other than the stated next-best alternative. The others are the irrelevance-defiers. In our context, the irrelevance assumption means that if crossing the admission cutoff to a given field does not make an individual choose that field, it should not affect her choice of other fields either.

In Section 3, we use this stratification of the population to characterize the bias in the IV estimands that may arise in the presence of next-best defiers, or irrelevance defiers, or both. It is useful to observe that the bias due to each type of defier has a product structure: It depends on the number of defiers compared to compliers, multiplied by the difference between compliers and defiers in the average payoff to choosing one type of education compared to another. Thus, there will be zero bias if there either are no defiers or if the average payoff to choosing one type of education compared to another is the same for defiers and compliers. Furthermore, the bias becomes large only if there are many defiers relative to compliers *and* there are large differences in the payoff between compliers and defiers.

In Section 4, we show that the shares of next-best and irrelevance defiers can be bounded, but not point identified. We derive sharp bounds – which are nontrivial – and, thus, provides testable implications of the additional assumptions of Kirkeboen et al. (2016). We show that these results have implications for the recent work of Nibbering et al. (2022) who propose an algorithm which aggregate fields into clusters based on estimated first-stage coefficients. The motivation for their approach is to avoid bias from irrelevance and next-best defiers. We show that their approach requires point identification of the shares of next-best and irrelevance defiers, and that it may produce biased estimates even if effects are constant across individuals (in contrast to standard 2SLS).

The last three sections of the paper take the theoretical results discussed above to the data by comparing payoff estimates for two countries, Denmark and Norway. These are two geographically and culturally close open-economies with very comparable educational institutions as well as similar tax, welfare and social benefits systems. It seems therefore natural to expect that payoffs will – at least to a degree – be aligned, and that differences can potentially be understood in light of violations of the assumptions underlying approach outlined above and detailed below.

In Section 5, we present the institutional background and data sources in Denmark and Norway. This section highlights the common institutional framework and data sources, documents how educational classifications and outcomes are harmonized across countries, and discusses differences that may be consequential for the analysis and results. In Section 6, we present the empirical specification that generates the payoff estimates for the two countries, following closely Kirkeboen et al. (2016). Two challenges that must be met when comparing estimates from two different populations are the reference population and measurement error. Section 6 therefore also defines the population of compliers that we use to anchor the estimates, and presents an error-in-variables approach that addresses bias arising from measurement error when comparing the noisy payoff estimates.

In Section 7, we present the estimation results. We first turn to the first-stages and, building on the results from Section 4, document that in both countries the violations of irrelevance or next-best are non-trivial and appear to be of similar magnitude but of a different nature. In Norway, there is clear evidence of violations of next-best, but little if any sign of violations of irrelevance; in Denmark, the two types of violations seem to be equally frequent. The accompanying second-stages (with earnings measured eight years after application) show that, on average, the estimated annual payoff to completing a field-of-study instead of the next-best is about 2,200 USD in Denmark, while in Norway the payoff estimates are substantially larger and around 22,000 USD. The payoffs in Norway also exhibit a higher variance than in Denmark, and overall we strongly reject that the payoffs are the same. Despite these differences in levels and variation the payoffs significantly co-vary, and we estimate a correlation coefficient of 0.65 for our population of interest.

This correlation substantially increases when we exclude the estimates with the most violations of the irrelevance and next-best assumptions. However, violations of irrelevance and next-best do not appear to explain the lower level and variation of the payoffs in Denmark compared to Norway. Additional exploratory analyses show that these across country differences are mostly driven by heterogeneity in next-best fields, and can partly be explained by differences in selectivity (as measured by students' high school test scores).

2 IV with multiple unordered treatments

2.1 Models and assumptions

We assume individuals choose between three mutually exclusive and collectively exhaustive alternatives $d \in \{0,1,2\}$. To fix ideas we envision these as enrolling in three different fields of study. We suppress the individual index and abstract from control variables. We want to interpret IV estimates of the equation

$$y = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \varepsilon \tag{1}$$

where y is an observed outcome such as earnings, and $d_j \equiv \mathbb{1}_{[d=j]}$ is a treatment indicator. Without loss of generality we choose field 0 as reference field, so that β_1^{IV} (β_2^{IV}) is the payoff from choosing field 1 (2) over field 0.

We suppose individuals are randomly assigned to one of three mutually exclusive and collectively exhaustive groups $Z \in \{0,1,2\}$ and let $z_j = \mathbb{1}_{[Z=j]}$ be an indicator variable that equals 1 if an individual is assigned to group j and 0 otherwise. The indicator z_j can be thought of as an instrument shifting the costs or benefits of choosing field j. For each individual, this gives three potential field choices d^z and nine potential outcomes $y^{d,z}$. We let $\mathbf{d}$ denote the column vector of treatment indicators and $\mathbf{z}$ the column vector of instruments. We define $d_j^z \equiv \mathbb{1}_{[d^z=j]}$ to be an indicator variable that tells us whether an individual would choose field j for a given value of Z.

As in the analysis of binary treatments in Imbens and Angrist (1994), we allow for heterogeneous effects and assume that each instrument satisfies the following assumptions:

Assumption 1. *IV Assumptions*

(a) **Exclusion:** $y^{d,z} = y^d$ *for all* d,z

(b) **Independence:** $y^d, d^z \perp Z$ *for all* d,z

(c) **Rank:** $E[\mathbf{z}\mathbf{d}^\top]$ *has full rank*

(d) **Monotonicity:** $d_k^k \geq d_k^{k'}$ *for each field assignment pair* k,k'

Given our notation and assumptions, we can link the observed and potential outcomes and choices as follows,

$$y = y^0 d_0 + y^1 d_1 + y^2 d_2 \tag{2}$$

$$d_j = d_j^0 z_0 + d_j^1 z_1 + d_j^2 z_2 \qquad \text{for } j = 0,1,2 \tag{3}$$

Table 1. Taxonomy of complier and defier groups with field 0 as the stated next-best alternative

Group		Potential Field Choice			Characteristics
		d^0	d^1	d^2	
Instrument 1					
- Compliers	C_1	0	1		$d_1^1 - d_1^0 = 1 \ \wedge \ d_2^1 = d_2^0 = \ 0$
- Irrelevance Defiers	ID_1	0	2		$d_1^1 = d_1^0 = 0 \ \wedge \ d_2^1 - d_2^0 = \ 1$
- Next-best Defiers	ND_1	2	1		$d_1^1 - d_1^0 = 1 \ \wedge \ d_2^1 - d_2^0 = -1$
Instrument 2					
- Compliers	C_2	0		2	$d_2^2 - d_2^0 = 1 \ \wedge \ d_1^2 = d_1^0 = \ 0$
- Irrelevance Defiers	ID_2	0		1	$d_2^2 = d_2^0 = 0 \ \wedge \ d_1^2 - d_1^0 = \ 1$
- Next-best Defiers	ND_2	1		2	$d_2^2 - d_2^0 = 1 \ \wedge \ d_1^2 - d_1^0 = -1$

Note: The table characterizes compliers, irrelevance defiers and next-best defiers based on their potential treatments.

These equations represent a model with multiple unordered treatment that permits unrestricted unobserved heterogeneity in treatment effects. Extending the model (and our theoretical results) to more than three choice alternatives is straightforward.

2.2 Principal strata

In Table 1 we invoke assumptions 1(a)–1(d) and characterize the principal strata, that is, the groups of individuals defined by how their potential field choices depend on the instrument. The table considers the (sub)population with field 0 as the stated next-best alternative. For brevity, it does not include always takers of field 1 (2) (those who chose field 1 (2) irrespective of instrument value) and never takers of field 1 (2) (those who choose field 2 and 0 (1 and 0) irrespective of instrument value).

As shown in the table, there are two types of compliers, C_1 and C_2. The C_1 (C_2) compliers are individuals who choose field 1 (2) when the instrument takes value 1 (2), and the reference field 0 when the instrument takes the value 0. In addition, there are four types of defiers, irrelevance and next-best defiers of instruments 1 and 2. Irrelevance defiers ID_1 (ID_2) are individuals who choose field 2 (1) when the instrument takes value 1 (2) while choosing field 0 if the instrument takes value 0. Next-best defiers ND_1 (ND_2) are individuals who choose field 2 (1) when the instrument takes value 0 while choosing field 1 (2) if the instrument takes value 1 (2).

2.3 Identification result

Kirkeboen et al. (2016) suggest the following assumptions on the groups in Table 1 to

obtain identification:[5]

Assumption 2. *Auxiliary Assumptions*

(a) **Irrelevance:** $d_k^k - d_k^0 = 0 \implies d_{k'}^k = d_{k'}^0$ *for all pairs* k, k'

(b) **Next-best:** *We are able to condition on* $d_1^0 = d_2^0 = 0$ *i.e.* $d_0^0 = 1$.

The irrelevance condition assumes that if changing z from 0 to 1 (2) does not induce an individual to choose treatment 1 (2), then it does not make her choose treatment 2 (1) either. In our context, for example, this assumption means that if crossing the admission cutoff to field 1 does not make an individual choose field 1, it does not make her choose field 2 either. The next-best alternative condition is effectively assuming that individuals' stated next-best alternative is their actual next-best alternative. The following lemma is immediate from these two assumptions:

Lemma 1. *Suppose Assumptions 1–2 hold. Then* $\beta_1^{IV}, \beta_2^{IV}$ *have a causal interpretation as positively weighted averages of treatment effects for compliers, and*

$$\beta_1^{IV} = \mathbb{E}[y^1 - y^0 \mid C_1]$$
$$\beta_2^{IV} = \mathbb{E}[y^2 - y^0 \mid C_2]$$

Proof. For a proof, see Kirkeboen et al. (2016). □

The core of Lemma 1 is that the IV estimand of β_1 (β_2) can be given an interpretation as a local average treatment effect (LATE) of an instrument-induced shift from field 0 to field 1 (2) for compliers when irrelevance and next-best defiers are assumed away.

3 Interpretation of IV estimands if auxiliary assumptions fail

If Assumptions 2(a)–2(b) do not hold, the IV estimand of β_1 (β_2) does not have a causal interpretation as a positively weighted average of treatment effects of choosing field 1 (2) over field 0. In the following, we characterize the bias that will occur in this case, and discuss in which situations the bias will be large and small.

[5] Kirkeboen et al. (2016) are imprecise about whether auxiliary assumption 2(b) is imposed on everyone or only those individuals whose treatment status depends on the instrument. However, this is immaterial for their results, as well as ours. The reason is that always takers and never takers drop out of the IV estimand because their treatment status does not change with the instrument.

3.1 *Assuming only next-best*

The IV estimands of β_1 and β_2 can be decomposed into a LATE for compliers and a bias term using IV moment conditions. In particular, if only next-best holds, but not irrelevance, we get the following decomposition, as shown in Appendix A.

Proposition 1. *Suppose Assumptions 1(a)–1(d) and 2(b) hold. Then $\beta_1^{IV}, \beta_2^{IV}$ do not have a causal interpretation as positively weighted averages of treatment effects for compliers,*

$$\begin{aligned}\beta_1^{IV} = \underbrace{\mathbb{E}[y^1 - y^0 \mid C_1]}_{\text{A}} \quad &+ \quad \underbrace{\frac{P(ID_1)P(ID_2)}{W'}}_{\omega_1} \times \underbrace{(\mathbb{E}[y^1 - y^0 \mid C_1] - \mathbb{E}[y^1 - y^0 \mid ID_2])}_{\Delta_1} \\ &- \quad \underbrace{\frac{P(ID_1)P(C_2)}{W'}}_{\omega_2} \times \underbrace{(\mathbb{E}[y^2 - y^0 \mid C_2] - \mathbb{E}[y^2 - y^0 \mid ID_1])}_{\Delta_2}\end{aligned} \tag{4}$$

where $W' = P(C_1)P(C_2) - P(ID_1)P(ID_2)$ and the expression for β_2^{IV} follows by symmetry. A *is the complier LATE, ω_1 and ω_2 are defier group weights, and Δ_1 and Δ_2 are differences in the causal effects between compliers and irrelevance defiers.*

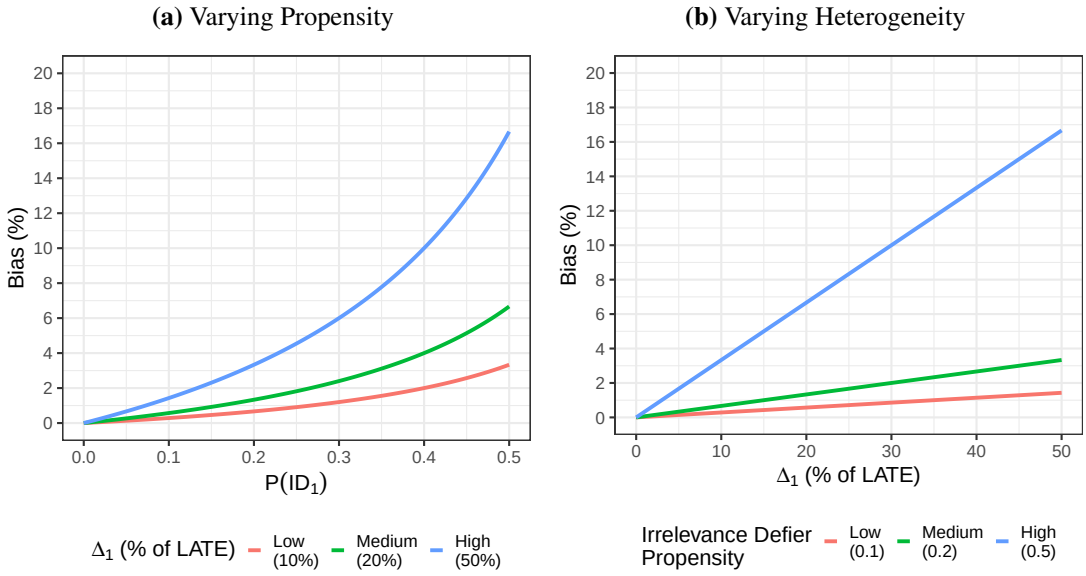
Proof. See appendix A. □

Imposing the constant effects assumption implies that the differences in the causal effects between defier groups (Δ_1, Δ_2) go to zero. In this case, β_1^{IV} (β_2^{IV}) would recover the causal effect, $\mathbb{E}[y^1 - y^0]$ ($\mathbb{E}[y^2 - y^0]$). Imposing irrelevance implies that the defier weights (ω_1, ω_2) go to zero. In this case, β_1^{IV} (β_2^{IV}) would recover the complier LATE, $\mathbb{E}[y^1 - y^0 \mid C_1]$ ($\mathbb{E}[y^2 - y^0 \mid C_2]$).

A central question for empirical researchers is when the bias in Proposition 1 is likely to be large. To answer this question, it is useful to observe that the two bias terms in equation 4 are the products of a difference in causal effects and a defier weight consisting of the product of the propensities of irrelevance defiers divided by the difference between complier and defier propensity products.

Note that as long as $P(C_1)P(C_2) > 2 \times P(ID_1)P(ID_2)$ the weight ω_1 is below 1. This will occur when there are many compliers relative to defiers. When the weight is below 1, the bias will always be smaller than the difference in causal effects. Due to the product structure ($\omega_j \times \Delta_j$) the bias due to violations of the irrelevance assumption will be very small when both ω_j and Δ_j are small. Conversely, in order for a large bias to occur, there needs to be both many defiers relative to compliers and a large difference in causal effects between the compliers and the irrelevance defiers.

We illustrate this with two examples. In both examples, we fix the LATE for compliers at \$1000. We focus on the first instrument, fixing the propensities of compliers and irrele-



Note: Panel (a) shows the bias from irrelevance defiers for different defier propensities. The red line assumes a difference in causal effects between compliers and defiers at 10% of the complier LATE, the green at 20% and the blue at 50%. Panel (b) shows the bias from irrelevance defiers for different levels of treatment effect heterogeneity. The red line assumes 10%, the green 20% and the blue 50% irrelevance defiers. The number of defiers and compliers for instrument 2 is fixed at 20% and 80%.

Figure 1. Bias from irrelevance defiers under different defier weights and levels of heterogeneity.

vance defiers of instrument 2 to $P(ID_2) = 0.2$ and $P(C_2) = 0.8$, and, for simplicity, assume no always takers or never takers for any of the instruments, such that $P(C_1) = 1 - P(ID_1)$.

In Figure 1a we show how the bias from the first term varies with the propensity of irrelevance defiers. We let the difference in causal effects between compliers and instrument 2-defiers be fixed at three different levels: 10%, 20% and 50% of the complier LATE. In Figure 1b we show the bias from the first term when varying the difference in causal effects between compliers and defiers. We let the propensity of irrelevance defiers be fixed at three different levels: low (0.1), middle (0.2), and high (0.5). The key take away is that the bias will be small even when there is a sizable number of defiers and a nontrivial difference in causal effects between the compliers and the defiers.

3.2 Assuming only irrelevance

If irrelevance holds, but next-best is not observed, we may decompose the IV estimand into a complier LATE and a bias term.

Proposition 2. *Suppose Assumptions 1(a)–1(d) and 2(a) hold. Then* $\beta_1^{IV}, \beta_2^{IV}$ *do not have*

a causal interpretation as positively weighted averages of treatment effects for compliers,

$$\begin{aligned}\beta_1^{IV} = \underbrace{\mathbb{E}[y^1 - y^0 \mid C_1]}_{\text{A}} &+ \underbrace{\frac{P(ND_1)P(C_2)}{\hat{W}}}_{\omega_3} \times (\underbrace{\mathbb{E}[y^1 - y^0 \mid ND_1] - \mathbb{E}[y^1 - y^0 \mid C_1]}_{\Delta_3}) \\ &- \underbrace{\frac{P(ND_1)P(C_2)}{\hat{W}}}_{\omega_4} \times (\underbrace{\mathbb{E}[y^2 - y^0 \mid ND_1] - \mathbb{E}[y^2 - y^0 \mid C_2]}_{\Delta_4}) \\ &+ \underbrace{\frac{P(ND_1)P(ND_2)}{\hat{W}}}_{\omega_5} \times (\underbrace{\mathbb{E}[y^1 - y^0 \mid ND_1] - \mathbb{E}[y^1 - y^0 \mid ND_2]}_{\Delta_5}) \\ &- \underbrace{\frac{P(ND_1)P(ND_2)}{\hat{W}}}_{\omega_6} \times (\underbrace{\mathbb{E}[y^2 - y^0 \mid ND_1] - \mathbb{E}[y^2 - y^0 \mid ND_2]}_{\Delta_6})\end{aligned} \tag{5}$$

where $\hat{W} = P(C_1)P(C_2) + P(C_1)P(ND_2) + P(ND_1)P(C_2)$ *and the expression for* β_2^{IV} *follows by symmetry.* A *is the complier LATE,* ω_3 *through* ω_6 *are defier group weights and* Δ_3 *through* Δ_6 *are differences in the causal effects between complier and defier groups.*

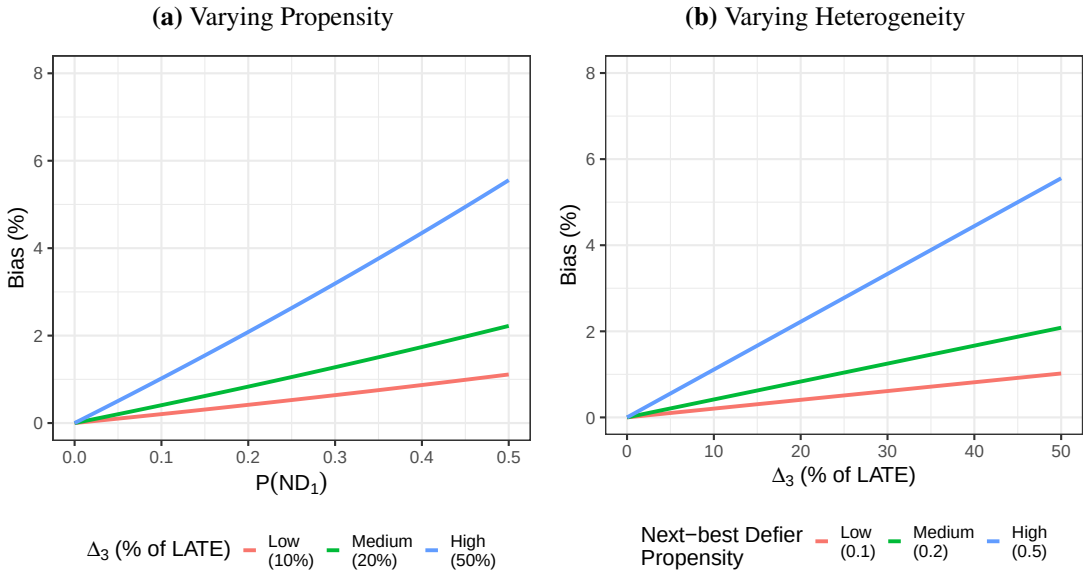
Proof. See appendix A. □

Imposing the constant effects assumption implies that the differences in causal effects between defier groups (Δ_3 through Δ_6) go to zero. In this case, β_1^{IV} (β_2^{IV}) would recover the causal effect, $\mathbb{E}[y^1 - y^0]$ ($\mathbb{E}[y^2 - y^0]$). Observing the next-best alternative implies that the defier weights (ω_3 through ω_6) go to zero. In this case, β_1^{IV} (β_2^{IV}) would recover the complier LATE, $\mathbb{E}[y^1 - y^0 \mid C_1]$ ($\mathbb{E}[y^2 - y^0 \mid C_2]$).

As in equation (4), the bias terms in equation (5) are the products of a difference in causal effects and a weight consisting of the product of the propensities of defiers divided by the sum of complier and defier propensity products.

Note that the weight in the first and second terms of equation (5) (ω_3, ω_4) are below 1, but that the weights for the two latter terms (ω_5, ω_6) can be above 1 if $P(ND_1)P(ND_2) > P(C_1)P(C_2) + P(C_1)P(ND_2) + P(ND_1)P(C_2)$. When the weight is below 1, the bias from the term will always be smaller than the difference in causal effects. Due to the product structure ($\omega_j \times \Delta_j$) the bias due to violations of the next-best assumption will be very small when both ω_j and Δ_j are small. Conversely, in order for a large bias to occur, we need both many defiers relative to compliers and a large difference in causal effects between the different groups.

We keep the same numerical example as in Section 3.1 and focus on the term $\omega_3 \times \Delta_3$. In Figure 2a we show how the bias from this term varies with the propensity of next-best



Note: Panel (a) shows one term of the bias from next-best defiers for different defier propensities. The red line assumes a difference in causal effects between compliers and defiers at 10% of the complier LATE, the green at 20% and the blue at 50%. Panel (b) shows the bias from irrelevance defiers for different levels of treatment effect heterogeneity. The red line assumes 10%, the green 20% and the blue 50% irrelevance defiers. The number of defiers and compliers for instrument 2 is fixed at 20% and 80%.

Figure 2. Bias from next-best defiers under different defier weights and levels of heterogeneity.

defiers. We let the difference in causal effects between compliers and defiers be fixed at three different levels: at 10%, 20% and 50% of the complier LATE. In Figure 2b we show the bias when varying the difference in causal effects between compliers and defiers. We let the propensity of next-best defiers be fixed at three different levels: low (0.1), middle (0.2), and high (0.5). The key take away is as above that the bias will be small even when there is a sizable number of defiers and a nontrivial difference in causal effects between the compliers and the defiers.

3.3 Assuming neither irrelevance nor next-best

If one neither makes the irrelevance assumption nor the next-best assumption, the IV estimand becomes the sum of the complier LATE, all bias terms from Propositions 1 and 2, as well as a third set of interacted bias terms.

Proposition 3. *Suppose Assumptions 1(a)–1(d) holds. Then $\beta_1^{IV}, \beta_2^{IV}$ do not have a causal interpretation as positively weighted averages of treatment effects for compliers,*

$$\beta_1^{IV} = \underbrace{\mathbb{E}[y^1 - y^0 \mid C_1]}_{\text{A}} + \underbrace{\frac{P(ID_1)P(ID_2)}{\bar{W}}}_{\omega_1} \times \underbrace{(\mathbb{E}[y^1 - y^0 \mid C_1] - \mathbb{E}[y^1 - y^0 \mid ID_2])}_{\Delta_1} \quad (6)$$

$$
\begin{aligned}
&- \underbrace{\frac{P(ID_1)P(C_2)}{\bar{W}}}_{\omega_2} \times \underbrace{(\mathbb{E}[y^2 - y^0 \mid C_2] - \mathbb{E}[y^2 - y^0 \mid ID_1])}_{\Delta_2} \\
&+ \underbrace{\frac{P(ND_1)P(C_2)}{\bar{W}}}_{\omega_3} \times \underbrace{(\mathbb{E}[y^1 - y^0 \mid ND_1] - \mathbb{E}[y^1 - y^0 \mid C_1])}_{\Delta_3} \\
&- \underbrace{\frac{P(ND_1)P(C_2)}{\bar{W}}}_{\omega_4} \times \underbrace{(\mathbb{E}[y^2 - y^0 \mid ND_1] - \mathbb{E}[y^2 - y^0 \mid C_2])}_{\Delta_4} \\
&+ \underbrace{\frac{P(ND_1)P(ND_2)}{\bar{W}}}_{\omega_5} \times \underbrace{(\mathbb{E}[y^1 - y^0 \mid ND_1] - \mathbb{E}[y^1 - y^0 \mid ND_2])}_{\Delta_5} \\
&- \underbrace{\frac{P(ND_1)P(ND_2)}{\bar{W}}}_{\omega_6} \times \underbrace{(\mathbb{E}[y^2 - y^0 \mid ND_1] - \mathbb{E}[y^2 - y^0 \mid ND_2])}_{\Delta_6} \\
&- \underbrace{\frac{P(ND_1)P(ID_2)}{\bar{W}}}_{\omega_7} \times \underbrace{(\mathbb{E}[y^1 - y^0 \mid C_1] - \mathbb{E}[y^1 - y^0 \mid ID_2])}_{\Delta_7} \\
&+ \underbrace{\frac{P(ID_1)P(ND_2)}{\bar{W}}}_{\omega_8} \times \underbrace{(\mathbb{E}[y^1 - y^0 \mid ND_2] - \mathbb{E}[y^1 - y^0 \mid C_1])}_{\Delta_8} \\
&- \underbrace{\frac{P(ID_1)P(ND_2)}{\bar{W}}}_{\omega_9} \times \underbrace{(\mathbb{E}[y^2 - y^0 \mid ND_2] - \mathbb{E}[y^2 - y^0 \mid ID_1])}_{\Delta_9}
\end{aligned}
$$

where

$$
\begin{aligned}
\bar{W} \;=\; & P(C_1)P(C_2) + P(C_1)P(ND_2) + P(ND_1)P(C_2) \\
& + P(ND_1)P(ID_2) + P(ID_1)P(ND_2) - P(ID_1)P(ID_2)
\end{aligned}
$$

and the expression for β_2^{IV} follows by symmetry.

Proof. See appendix A. □

A is the complier LATE, ω_1 and ω_2 are defier weights which also occur when observing the next-best alternative, ω_3 through ω_6 are defier weights which also occur under irrelevance and ω_7 through ω_9 are defier weights which occur only when neither assumption holds. Δ_1, Δ_2 and Δ_7 are differences in the causal effects between irrelevance defiers and compliers, Δ_3, Δ_4 and Δ_8 are differences in the causal effects between next-best defiers and compliers, while Δ_5 and Δ_6 are differences in causal effects between next-best defiers for the two different instruments, and Δ_9 is the difference in the causal effects between next-best defiers of instrument 2 and irrelevance defiers of instrument 1.

Imposing the constant effects assumption implies that the differences in causal effects between defier groups (Δ_1 through Δ_9) go to zero. In this case, β_1^{IV} (β_2^{IV}) would recover the causal effect, $\mathbb{E}[y^1 - y^0]$ ($\mathbb{E}[y^2 - y^0]$). Imposing the next-best assumption yields the result from Proposition 1, as weights ω_3 through ω_9 go to zero. Imposing the irrelevance assumption yields Proposition 2, as weights ω_1, ω_2 and ω_7 through ω_9 go to zero. Imposing both irrelevance and observing the next-best alternative make all defier weights (ω_1 through ω_9) go to zero. Then β_1^{IV} (β_2^{IV}) would recover the complier LATE, $\mathbb{E}[y^1 - y^0 \mid C_1]$ ($\mathbb{E}[y^2 - y^0 \mid C_2]$).

Note that the bias in Proposition 3 is the sum of all bias terms from Propositions 2 and 1, in addition to three new bias terms (except for a different denominator of the weights). These are terms following from interactions between irrelevance and next-best defiers, and rely on both types of defiers being present and having differences in causal effects between each other and with the complier group. As a result, the bias will be small unless there are relatively many of *both* types of defiers and the causal effects are materially different between these groups and the compliers.

4 Testable implications and aggregation

4.1 How to test the auxiliary assumptions

The first stage equations for the IV estimates of equation (1) are given by:

$$d_1 = \alpha_1^0 + \alpha_1^1 z_1 + \alpha_1^2 z_2 + \nu_1 \tag{7}$$

$$d_2 = \alpha_2^0 + \alpha_2^1 z_1 + \alpha_2^2 z_2 + \nu_2 \tag{8}$$

We now examine if it is possible to devise a test of whether the auxiliary assumptions 2(a)–2(b) hold empirically. To do so, it is useful to characterize the quantities that the first stage coefficients recover:

Lemma 2. *Suppose Assumptions 1(a)–1(d) hold. Then*

$$\alpha_1^0 = P(AT_1) \equiv P(OT_2) + P(ND_2) \qquad \alpha_2^0 = P(AT_2) \equiv P(OT_1) + P(ND_1)$$
$$\alpha_1^1 = P(C_1) + P(ND_1) \qquad \alpha_2^2 = P(C_2) + P(ND_2)$$
$$\alpha_2^1 = P(ID_1) - P(ND_1) \qquad \alpha_1^2 = P(ID_2) - P(ND_2)$$

where AT_1(AT_2) are always-takers of field 1 (2) when z equals 0 or 1 (0 or 2), and OT_1 (OT_2) are global (for every value of the instrument) always takers of the other field 2 (1). See Appendix Table A1 for formal definitions of these instrument-specific strata.

Proof. See appendix B. □

This result paves the way for the main result on the testability of the irrelevance and next best assumptions:

Proposition 4. *Suppose Assumptions 1(a)–1(d) hold. Then $P(ID_1)$ and $P(ND_1)$ are partially identified.*

$$\begin{aligned} P(ND_1) &\in [\max\{0, -\alpha_2^1\}, \quad \min\{\alpha_1^1, \alpha_2^0\}] \\ P(ID_1) &\in [\max\{0, \quad \alpha_2^1\}, \max\{0, \alpha_2^1 + \min\{\alpha_1^1, \alpha_2^0\}\}] \end{aligned}$$

where results for $P(ID_2)$ and $P(ND_2)$ follow by symmetry.

Proof. See appendix B. □

The practical implication of Proposition 4 is that we cannot point identify the defier propensities without further assumptions. Yet, the assumptions are testable as the bounds will generally be nontrivial. Furthermore, if either assumption 2(a) or 2(b) is known to hold, the other assumption can be tested separately and $P(ID_1)$ or $P(ND_1)$ is point identified.

Corollary 1. *Suppose Assumptions 1(a)–1(d) and 2(b) hold. Then $P(ND_1) = P(ND_2) = 0$ and we can test whether assumption 2(a) (irrelevance) holds, as $\alpha_2^1 = P(ID_1)$ and $\alpha_1^2 = P(ID_2)$.*

Corollary 2. *Suppose Assumptions 1(a)–1(d) and 2(a) hold. Then $P(ID_1) = P(ID_2) = 0$ and we can test whether assumption 2(b) (next-best) holds, as $\alpha_2^1 = -P(ND_1)$ and $\alpha_1^2 = -P(ND_2)$.*

4.2 How aggregation may cause violations of the exclusion restriction

Nibbering et al. (2022) propose an algorithm which aggregates fields into clusters based on estimated first-stage coefficients. The motivation for their approach is to avoid bias from irrelevance and next-best defiers. Before discussing their approach, it is important to observe that the resulting IV estimates between such clusters will, at best, identify a positively weighted average of the causal effects of choosing one field versus a linear combination of the other fields, for example, the effects of choosing field 1 versus field 0 or 2. Hence, this approach involves moving the goalpost from clearly defined field contrasts that govern individuals' educational investments to clusters of different fields. In the discussion below, we accept at faith that such contrasts are parameters of interest.

Table 2. Four Possible Clustering Scenarios.

Scenario	Conditions		Clusters			Implied Restrictions on Defiers
	α_2^1	α_1^2	S_0	S_1	S_2	
Control Clustering	<0	$=0$	$\{0,2\}$	$\{1\}$		$P(ND_1) > P(ID_1) \geq 0 \wedge P(ID_2) = P(ND_2) \geq 0$
	<0	>0				$P(ND_1) > P(ID_1) \geq 0 \wedge P(ID_2) > P(ND_2) \geq 0$
	$=0$	<0	$\{0,1\}$	$\{2\}$		$P(ID_1) = P(ND_1) \geq 0 \wedge P(ND_2) > P(ID_2) \geq 0$
	>0	<0				$P(ID_1) > P(ND_1) \geq 0 \wedge P(ND_2) > P(ID_2) \geq 0$
Treatment Clustering	>0	$=0$	$\{0\}$	$\{1,2\}$		$P(ID_1) > P(ND_1) \geq 0 \wedge P(ID_2) = P(ND_2) \geq 0$
	$=0$	>0				$P(ID_1) = P(ND_1) \geq 0 \wedge P(ID_2) > P(ND_2) \geq 0$
	>0	>0				$P(ID_1) > P(ND_1) \geq 0 \wedge P(ID_2) > P(ND_2) \geq 0$
No Clustering	$=0$	$=0$	$\{0\}$	$\{1\}$	$\{2\}$	$P(ID_1) = P(ND_1) \geq 0 \wedge P(ID_2) = P(ND_2) \geq 0$
Undefined*	<0	<0	$\{0,2\}$/ $\{0,1\}$	$\{1\}$/ $\{2\}$		$P(ND_1) > P(ID_1) \geq 0 \wedge P(ND_2) > P(ID_2) \geq 0$

Note: The table shows different clusterings ensuing from the algorithm proposed by Nibbering et al. (2022) and their implied restrictions on defiers. The algorithm tests the null hypothesis of coefficients being zero. The conditions in columns two and three specify which estimates must be observed for the clustering to be chosen, where >0 (<0) indicate rejecting the null and observing a positive (negative) coefficient, while "$=0$" indicates not being able to reject.
*It is unclear what Nibbering et al. (2022) do when both coefficients are negative. In that case, the ordering of the coefficients will matter.

4.2.1 Bias From Exclusion Violation We continue to consider the situation with three fields, discussed above. The algorithm takes as a starting point all individuals with a certain reported next-best alternative (in our case taken to be 0), and test the hypothesis that the off-diagonal coefficients, α_2^1 and α_1^2, are zero. If this hypothesis is rejected, the sign of the coefficient is evaluated and the treatments are clustered according to the rules laid out in Table 2. For example, if α_2^1 is negative and α_1^2 is either zero or positive, fields 0 and 2 become the control cluster and field 1 the treatment cluster. Conversely, if α_2^1 is either zero or positive and α_1^2 is negative, fields 0 and 1 become the control cluster and field 2 the treatment cluster.

After performing the clustering based algorithm, Nibbering et al. (2022) estimate cluster treatment effects: Let $\tilde{d}(d) = \mathbb{1}_{[d \in S_1]}$ be the binary cluster treatment indicator and $\tilde{z}(Z) = \mathbb{1}_{[Z=d \in S_1]}$ the cluster instrument indicator. The no clustering-scenario is equivalent to the field level. In the two other scenarios (control clustering or treatment clustering) we consider IV estimates of the equation

$$y = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{d} + \varepsilon$$

where the first stage is

$$\tilde{d} = \pi_0 + \pi_{1,0} \tilde{z} + \nu$$

and $\pi_{1,0}$ is the first stage coefficient. Observed and potential outcomes and choices are

linked as

$$y = \tilde{y}^0(1-\tilde{d}) + \tilde{y}^1\tilde{d} \tag{9}$$

$$\tilde{d} = \tilde{d}^0 + (\tilde{d}^1 - \tilde{d}^0)\tilde{z} \tag{10}$$

where $\tilde{d}^j \equiv \mathbb{1}_{[\tilde{d}^j=1]}$ denotes the cluster-level potential treatment and $\tilde{y}^j$ is the potential outcome in cluster j. In Appendix C we show that this IV estimand does not, under Assumptions 1(a)–1(d), have a causal interpretation as a positively weighted average of treatment effects for the cluster complier groups. This result is summarized in Proposition 5.

Proposition 5. *Suppose Assumptions 1(a)–1(d) hold.*

(a) *Under control clustering, $\tilde{\beta}_1^{IV}$ does not have a causal interpretation as a positively weighted average of treatment effects for the cluster complier group. If the clustering is $S_1 = \{1\}$ and $S_0 = \{2,0\}$, we have*

$$\tilde{\beta}_{1,0}^{IV} = \underbrace{\frac{P(C_1 \cup ND_2)}{\pi_{1,0}}\mathbb{E}[y^1 - y^0 \mid C_1 \cup ND_2] + \frac{P(C_2 \cup ND_1)}{\pi_{1,0}}\mathbb{E}[y^1 - y^2 \mid C_2 \cup ND_1]}_{A}$$
$$+ \underbrace{\frac{P(ID_1)}{\pi_{1,0}}}_{\tilde{\omega}_1}\underbrace{\mathbb{E}[y^2 - y^0 \mid ID_1]}_{\tilde{\Delta}_1} - \underbrace{\frac{P(ID_2)}{\pi_{1,0}}}_{\tilde{\omega}_2}\underbrace{\mathbb{E}[y^2 - y^0 \mid ID_2]}_{\tilde{\Delta}_1}$$

where $\pi_{1,0} = P(C_1 \cup C_2 \cup ND_1 \cup ND_2)$. A is a positively weighted average of cluster complier LATEs, $\tilde{\omega}_1$ and $\tilde{\omega}_2$ are defier group weights, and $\tilde{\Delta}_1$ and $\tilde{\Delta}_2$ are differences in potential outcomes for irrelevance defiers in cluster S_0, i.e. never takers of the clustered treatment. The result for the clustering $S_1 = \{2\}$ and $S_0 = \{1,0\}$ is symmetric.

(b) *Under treatment clustering, $\tilde{\beta}_1^{IV}$ does not have a causal interpretation as a positively weighted average of treatment effects for the cluster complier group. We have*

$$\tilde{\beta}_{1,0}^{IV} = \underbrace{\frac{P(C_1 \cup ID_2)}{\pi_{1,0}}\mathbb{E}[y^1 - y^0 \mid C_1 \cup ID_2] + \frac{P(C_2 \cup ID_1)}{\pi_{1,0}}\mathbb{E}[y^2 - y^0 \mid C_2 \cup ID_1]}_{A}$$
$$+ \underbrace{\frac{P(ND_1)}{\pi_{1,0}}}_{\tilde{\omega}_3}\underbrace{\mathbb{E}[y^1 - y^2 \mid ND_1]}_{\tilde{\Delta}_3} - \underbrace{\frac{P(ND_2)}{\pi_{1,0}}}_{\tilde{\omega}_4}\underbrace{\mathbb{E}[y^1 - y^2 \mid ND_2]}_{\tilde{\Delta}_4}$$

where $\pi_{1,0} = P(C_1 \cup C_2 \cup ID_1 \cup ID_2)$. A *is a positively weighted average of cluster complier LATEs,* $\tilde{\omega}_3$ *and* $\tilde{\omega}_4$ *are defier group weights, and* $\tilde{\Delta}_3$ *and* $\tilde{\Delta}_4$ *are differences in potential outcomes for irrelevance defiers in cluster* S_1*, i.e. always takers of the clustered treatment.*

Proof. See Appendix C. □

Imposing the irrelevance assumption under control clustering implies that the defier weights $(\tilde{\omega}_1, \tilde{\omega}_2)$ go to zero. In this case, $\tilde{\beta}^{IV}_{1,0}$ recovers a positively weighted average of the causal effect of choosing field 1 over 0 for compliers of instrument 1 and next-best defiers of instrument 2, and of choosing field 1 over 2 for compliers of instrument 2 and next-best defiers of instrument 1, weighted by the number of compliers and defiers. Under control clustering, this is the new parameter of interest.

Imposing the next-best assumption under treatment clustering implies that the defier weights $(\tilde{\omega}_3, \tilde{\omega}_4)$ go to zero. In this case, $\tilde{\beta}^{IV}_{1,0}$ recovers a positively weighted average of the causal effect of choosing field 1 over 0 for compliers of instrument 1 and irrelevance defiers of instrument 2, and of choosing field 2 over 0 for compliers of instrument 2 and irrelevance defiers of instrument 1, weighted by the number of compliers and defiers. Under treatment clustering, this is the new parameter of interest.

If neither irrelevance nor next-best assumptions hold, the IV estimand does not have a causal interpretation as a positively weighted average of treatment effects for the cluster complier group. The bias terms reflect that individuals may in response to changes in the cluster instrument be switching across fields in the treatment cluster and/or across fields in the control cluster. Such switches will generally involve changes in potential outcomes, yet no change in the cluster treatment status. Thus, the exclusion restriction at the cluster level will be violated. The reason for this bias is that the algorithm equates the sign of the off-diagonal coefficients with the presence and absence of irrelevance and nex-best defiers. As shown in Lemma 2, this is wrong. The off-diagonal coefficients tell us only if there are more or less next-best defiers than irrelevance defiers. One cannot in general use the sign of α^1_2 (α^2_1) to show that there are no irrelevance defiers of instrument 1 (2) if $\alpha^1_2 < 0$ $(\alpha^2_1 < 0)$ and no next-best defiers of instrument 1 (2) if $\alpha^1_2 > 0$ $(\alpha^2_1 > 0)$.

It is also important to observe that the constant effects assumption is not sufficient for $\tilde{\beta}^{IV}_{1,0}$ to recover a positively weighted average of treatment effects between clusters 0 and 1 and obtain a causal interpretation. This result is summarized in Proposition 6.[6]

Proposition 6. *Suppose Assumptions 1(a)–1(d) hold and we further assume constant treatment effects.*

[6]One exception to this negative result is the special case in which the number of defiers for each instrument happen to be equal, i.e. that $P(ID_1) = P(ID_2)$ under control clustering or $P(ND_1) = P(ND_2)$ under treatment clustering.

(a) *Under control clustering, $\tilde{\beta}_1^{IV}$ does not recover the causal effect. If the clustering is $S_1 = \{1\}$ and $S_0 = \{2,0\}$, we have*

$$\tilde{\beta}_{1,0}^{IV} = \underbrace{\frac{P(C_1 \cup ND_2)}{\pi_{1,0}}\mathbb{E}[y^1 - y^0] + \frac{P(C_2 \cup ND_1)}{\pi_{1,0}}\mathbb{E}[y^1 - y^2]}_{A} + \underbrace{\frac{P(ID_1) - P(ID_2)}{\pi_{1,0}}}_{\dot{\omega}_1} \underbrace{\mathbb{E}[y^2 - y^0]}_{\dot{\Delta}_1}$$

where $\pi_{1,0} = P(C_1 \cup C_2 \cup ND_1 \cup ND_2)$. A *is a positively weighted average of the causal effects of choosing field 1 over 0 and of choosing field 1 over 2, $\dot{\omega}_1$ is a difference between defier group weights, and $\dot{\Delta}_1$ is the difference in potential outcomes for irrelevance defiers in cluster S_0, i.e. never takers of the clustered treatment. The result for the clustering $S_1 = \{2\}$ and $S_0 = \{1,0\}$ is symmetric.*

(b) *Under treatment clustering, $\tilde{\beta}_1^{IV}$ does not recover the causal effect. We have*

$$\tilde{\beta}_{1,0}^{IV} = \underbrace{\frac{P(C_1 \cup ID_2)}{\pi_{1,0}}\mathbb{E}[y^1 - y^0] + \frac{P(C_2 \cup ID_1)}{\pi_{1,0}}\mathbb{E}[y^2 - y^0]}_{A} + \underbrace{\frac{P(ND_1) - P(ND_2)}{\pi_{1,0}}}_{\dot{\omega}_2} \underbrace{\mathbb{E}[y^1 - y^2]}_{\dot{\Delta}_2}$$

where $\pi_{1,0} = P(C_1 \cup C_2 \cup ID_1 \cup ID_2)$. A *is a positively weighted average of the causal effects of choosing field 1 over 0 and of choosing field 2 over 0, $\dot{\omega}_2$ is a difference between defier group weights, and $\dot{\Delta}_2$ is the difference in potential outcomes for irrelevance defiers in cluster S_1, i.e. always takers of the clustered treatment.*

Proof. The constant effects assumption reduces all conditional expectations to unconditional expectations, i.e. $\mathbb{E}[y^j - y^k \mid G] = \mathbb{E}[y^j - y^k]$ for any group G and any combination of fields j,k. The result is immediate. □

In contrast, the approach of Kirkeboen et al. (2016) recovers the causal effect under the constant effects assumption. This shows that the clustering method relies on different, not weaker assumptions than Kirkeboen et al. (2016).

The following auxiliary exclusion restriction can be made to obtain identification under the clustering approach.

Assumption 3. *Cluster Exclusion Assumptions*

(a) **Control Cluster Exclusion:** $\tilde{d}^1 = \tilde{d}^0 = 0 \implies \tilde{y}^{0,1} = \tilde{y}^{0,0}$

(b) **Treatment Cluster Exclusion:** $\tilde{d}^1 = \tilde{d}^0 = 1 \implies \tilde{y}^{1,1} = \tilde{y}^{1,0}$

Assumptions 3(a) and 3(b) ensure that the bias from switchers within clusters (irrelevance defiers under control clustering and next-best defiers under treatment clustering) disappear, irrespective of the number of switchers. These assumptions are homogeneity restrictions on potential outcomes across different fields, and, thus, difficult to justify. Nevertheless, if one is willing to invoke Assumptions 3(a) and 3(b), one may obtain the following identification result:

Proposition 7. *Under control clustering, suppose Assumptions 1(a)–1(d) and 3(a) hold.* $\tilde{\beta}_1^{IV}$ *has a causal interpretation as the positively weighted average of treatment effects for cluster compliers. If the clustering is* $S_1 = \{1\}$ *and* $S_0 = \{2,0\}$*, we have*

$$\tilde{\beta}_{1,0}^{IV} = \frac{P(C_1 \cup ND_2)}{\pi_{1,0}} \mathbb{E}[y^1 - y^0 \mid C_1 \cup ND_2] + \frac{P(C_2 \cup ND_1)}{\pi_{1,0}} \mathbb{E}[y^1 - y^2 \mid C_2 \cup ND_1]$$

where $\pi_{1,0} = P(C_1 \cup C_2 \cup ND_1 \cup ND_2)$*. The result for clustering* $S_1 = \{2\}$ *and* $S_0 = \{1,0\}$ *is symmetric.*
Under treatment clustering, suppose Assumptions 1(a)–1(d) and 3(b) hold. $\tilde{\beta}_1^{IV}$ *has a causal interpretation as a positively weighted average of treatment effects for cluster compliers, and*

$$\tilde{\beta}_{1,0}^{IV} = \frac{P(C_1 \cup ID_1)}{\pi_{1,0}} \mathbb{E}[y^1 - y^0 \mid C_1 \cup ID_1] + \frac{P(C_2 \cup ID_1)}{\pi_{1,0}} \mathbb{E}[y^2 - y^0 \mid C_2 \cup ID_1]$$

where $\pi_{1,0} = P(C_1 \cup C_2 \cup ID_1 \cup ID_2)$*.*

Proof. Assumption 3(a) (3(b)) eliminates the bias terms in the results from Proposition 5 by letting $\tilde{\Delta}_1, \tilde{\Delta}_2$ ($\tilde{\Delta}_3, \tilde{\Delta}_4$) go to zero. The result is immediate. □

5 Empirical analysis

Guided and motivated by the formal results above, we now turn to the empirical analysis of the payoffs to field of study in Norway and Denmark.

5.1 Institutional settings

The Danish and Norwegian post-secondary education systems are similar in many respects. Their post-secondary education sectors consist of public universities and a larger number of public and private university colleges. The vast majority of students attend

a public institution, and even the private institutions are publicly funded and regulated. Universities all offer a wide selection of fields. By comparison, the university colleges rarely offer fields like Law, Medicine, Science, or Technology, but tend to offer professional degrees in fields like Engineering, Health, Business, and Teaching. Obtaining a post-secondary degree normally requires three to five years; there are no tuition fees; most students receive financial support (in the form of grants/loans) from the state.

The admission process is centralized in both countries. Applications are submitted to a central organization that handles the admission process to universities and university colleges. An applicant ranks programs (up to 15 in Norway and 8 in Denmark), each defined by a detailed field and an institution. The number of slots for each program is effectively determined by each country's ministry of education. For many programs, demand exceeds supply. Most slots in programs with excess demand are filled based on an application score derived from high school GPA. Offers are determined by the applicants' application score: the highest ranked applicant receives an offer for her preferred program; the second highest applicant receives an offer for her highest ranked program among the remaining programs; and so on. This is repeated until either slots run out, or applicants run out. This allocation mechanism corresponds to a so-called serial dictatorship, which is both Pareto efficient and strategy-proof (Svensson, 1999) and should therefore elicit the applicants' true ranking of fields at the time of application.[7] If students want to change field or institution, they usually need to participate in next year's admission process on equal terms with other applicants.[8]

For both countries, the exact thresholds are unpredictable at the time of application. They are not published until after the allocation process, and variation in thresholds over time is considerable. For programs with excess demand, the admission process implies that applicants scoring above a certain threshold are much more likely to receive an offer for a program they prefer compared to applicants with the same program preferences but marginally lower application score. This gives rise to credible instruments from discontinuities that effectively randomize applicants near admission cutoffs into different programs.

As explained in greater detail in Kirkeboen et al. (2016), the instruments are defined around local course rankings on students' application lists. These local rankings define

[7] A possible threat to strategy-proofness is the truncation of the application list (at 15 programs in Norway and 8 in Denmark) which might induce individuals to list a safe option as their last choice. However, this is likely unimportant in practice, as less than 0.1 percent of Norwegian applicants are offered their 15th choice, and less than 1 percent of Danish applicants list eight programs.

[8] Most programs in Denmark also have a standby (waiting) list and the GPA threshold for the standby list is typically a little lower than the main threshold. On the application form, applicants can choose whether to apply for the standby list. Applicants admitted to the standby list are guaranteed a study place the following year, but they are not considered for any of the lower-ranked programs on their application. Appendix E provides a more detailed discussion.

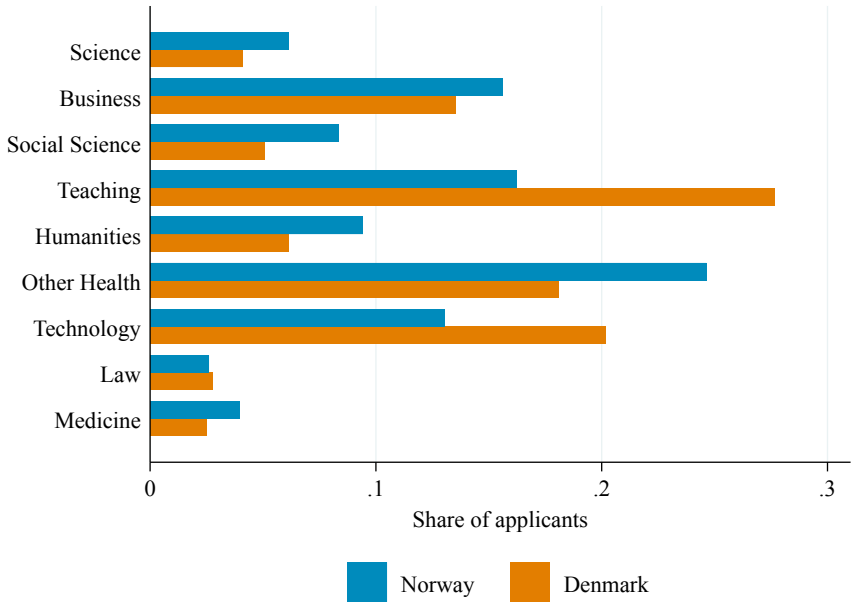
the "preferred" and "next-best" alternatives. For example, consider two fields, A and B, with A having a higher admission cutoff than B. Consider students who rank A just above B and have an application score that is either just below or just above the admission cutoff to A. These students will have A as the preferred field and B as the next-best field, no matter if A and B are ranked at the top, in the middle or at the bottom of the list. In other words, what matters for the relevance of instrument and the definition of preferred and next-best is the local ranking at which individuals are shifted in or out of a program because the application score is slightly above or below the relevant admission cutoff.

5.2 Data and descriptive statistics

For each country, we combine several sources of administrative data. For Norway, we use data for all applications to post-secondary education for the years 1998–2004. For Denmark, we use data for all applications to post-secondary education for the years 1994–2002. For both countries we retain the individuals' first observed application and exclude those who had a post-secondary degree at the time of application. We link these applicants to the population register and other registers to obtain background information, information on completed field, and annual earnings. In our main analysis we use data on treatment (completed field) and outcome (annual earnings) eight years after application as in Kirkeboen et al. (2016), and restrict the sample to those who have completed a field within eight years from application. The measure of earnings includes wage income, income from self-employment, and transfers that replace such income like short-term sickness pay and paid parental leave (but excludes unemployment benefits). Earnings are deflated using the CPI with 2011 as base year and converted to 1,000s of US dollars using the average exchange rates for the years 2010–2016 (6.5 Norwegian and 5.9 Danish crowns per US dollar).

We aggregate detailed fields into nine broad fields of study. We essentially follow the same classification of fields as in Kirkeboen et al. (2016). The only difference is that Technology now covers the integrated and more vocational/professional short and long cycle degrees at university colleges and universities and consist mostly of computer science and engineering degrees. Science corresponds to more open-ended bachelor programs in different sciences such as physics, biology and mathematics, as well as agriculture, forestry and aquaculture. For the main analysis, we retain all applicants who applied to at least two broad fields, where the most preferred field has an admission cutoff. If an applicant applied to several programs within her preferred broad field we use the lowest program cutoff as the effective cutoff to the preferred field.

While the full sample of applicants is of comparable size for the two countries, the final estimation sample is smaller in Denmark than in Norway, primarily because of fewer



Note: Figure shows number of applicants by completed field eight years after applying (conditional on having completed a field).

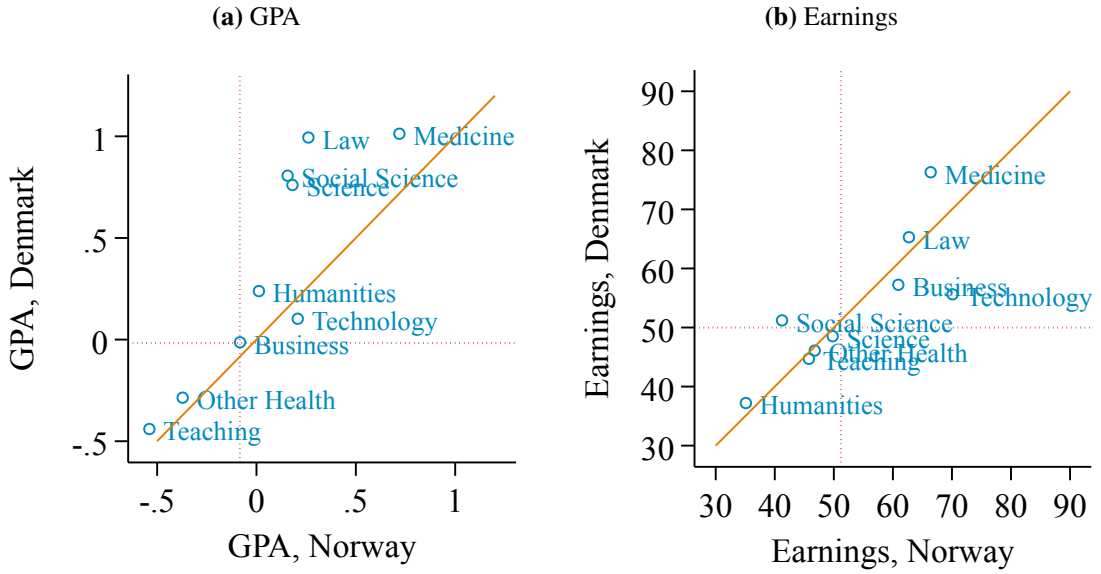
Figure 3. Distribution of applicants' completed field

fields with admission restrictions in Denmark. Figure 3 shows the distributions of completed field among applicants in Norway and Denmark eight years after applying. While the distributions are similar, there are some notable differences. The share of applicants completing teaching is substantially higher in Denmark than in Norway, as is the share of applicants having completed a degree with Technology. On the flip side the shares in Science, Social Science and Humanities are larger in Norway than in Denmark.[9]

As an indicator of relative selectivity, we standardize high school GPA within country and show in sub-graph (a) of Figure 4 the average standardized GPA by field and country. In both countries average GPA is relatively low in Teaching and Other Health. Average GPA is very high for Medicine in both countries, but Law, Social Science and Science are nearly equally selective as Medicine in Denmark.

Sub-graph (b) of Figure 4 compares earnings by field across country. Average earnings levels, indicated by the red dotted lines, are very similar in the two countries. Earnings in Medicine and Social Science are higher in Denmark consistent with their higher selectivity, but the same is not observed for Law and Science. Earnings are higher in Norway for

[9]Some of the cross-country differences may be due to differences in the classification of specific fields into the nine broad fields. For instance, one reason why the share having completed Teaching is large in Denmark is that all individuals having completed a bachelor's degree in social education are included in Teaching regardless of the specialization (e.g., kindergarten teacher, nursery teacher, nursery nurse, child and youth worker, support worker), while some of the specializations could alternatively be classified as Other Health if they were observed as separate educations.



Note: Figure shows applicant-weighted average GPA and earnings. GPA is demeaned and standardized within country. Earnings are CPI adjusted and converted to USD using fixed exchange rates (see text for details) and observed eight years after applying.

Figure 4. Applicants' GPA and earnings in Norway and Denmark, average by country and completed field of study

Technology. In both countries, earnings are particularly low for Humanities. However, it is important to note that earnings are measured eight years after application, which is very early in the career, especially for those choosing longer programs or programs characterized by a more difficult school-to-work transition. We examine the importance of this issue in a specification check that uses earnings measured later in the working life as the outcome variable.

In Appendix Figures A1 and A2 we present results similar to Figures 3 and 4, but not restricted to applicants. For Norway, data for Figures A1 and A2 consist of everybody born 1979–1983 such that we have application data for the years they are aged 19–21. Similarly, for Denmark the population sample consists of the cohorts born 1975–1981. Completed field and earnings are measured at age 28. The results for these broader populations are similar to the results for applicants in Figures 3 and 4.

6 How we estimate and compare payoffs

6.1 2SLS specification

The identification results in Sections 2 and 3 motivate and guide the specification of the empirical model. We consider the following system of equations separately for individu-

als with next-best field l (in the local field ranking):

$$y = \sum_{j \neq l} \beta_{jl} d_j + x' \gamma_l + \lambda_l^k + \varepsilon_l \tag{11}$$

$$d_j = \sum_{k \neq l} \alpha_{jl}^k z_k + x' \psi_{jl} + \eta_{jl}^k + u_{jl} \tag{12}$$

where (11) is the second stage equation, and (12) are the first-stage equations, one for each field. In these equations, j denotes the completed field, l denotes the stated next best alternative (in the local field ranking), and k denotes the preferred field (in the local field ranking). The l index is necessary in these equations, since we are now considering all possible preferred and next best fields (and not only focusing on field 0 as the stated next best alternative, as we did in the simple example in equations (2)-(3)).

The instruments z_k in (12) are the predicted offers for field k, and z_k is therefore equal to one if k is the individual's preferred field and her application score exceeds the admission cutoff for field k and zero otherwise. We therefore have as many binary instruments as treatments (one z_k for each completed field dummy d_j), and for a given individual at most one of the instruments z_k can equal 1 (namely the one of her preferred field in the local field ranking).

Our estimation approach exploits the fuzzy regression discontinuity design implicit in the admission process described above, where individuals with application scores above the cutoff are more likely to receive an offer for their preferred field. Although the identification in this setup is ultimately local, we use 2SLS because our sample sizes do not allow for local non-parametric estimation. While the model laid out above abstracted from any control variables, we now need to include certain covariates to ensure the exogeneity of our instruments.

First, all equations include controls for the running variable. While our baseline specification controls for the application score linearly on each side of the admission cutoff, Kirkeboen et al. (2016) reported results from several specification checks, all of which support our main findings. Second, we control for individuals' preferences by adding fixed effects for preferring field k and having l as the next-best field (in the local field ranking): λ_l^k and η_{jl}^k. To gain precision, we estimate the system of equations (11)–(12) jointly for all completed and next-best fields, allowing for separate intercepts for preferred field and for next-best field by completed field (i.e. $\lambda_l^k = \mu^k + \theta_j$ and $\eta_{jl}^k = \tau_j^k + \sigma_j^k$). In a robustness check, Kirkeboen et al. (2016) show that their estimates are robust to allowing for separate intercepts for every interaction between preferred and next-best field. Finally, to reduce residual variance we also add controls for gender, cohort and age at application, which are pre-determined.

From the resulting 2SLS estimation of equations (11)–(12) across all next-best fields, we obtain a matrix of the payoffs to field j compared to k for those who prefer j and have k as next-best field. In our baseline specification of the fields, we have 9 completed fields (j), 9 possible preferred fields/instruments (k), 8 possible next-best fields (l).[10] Because preferred field can never be the same as the next-best alternative, we get 576 (and not 648) unique first stage coefficients, α_{jl}^{k}. Because $\sum d_j = 1$ for each applicant, creating a within-applicant correlation between different d_j, we allow the residuals u_{jl} to be clustered within applicant.

6.2 *Comparing payoff estimates*

We want to compare payoffs to field of study across two different populations:

$$(\beta_{jl}^{DK} - \overline{\beta}^{NO}) = a_0 + a_1(\beta_{jl}^{NO} - \overline{\beta}^{NO}) + e_{jl} \tag{13}$$

where we have re-centered the payoffs relative to the average Norwegian payoffs for interpretational convenience: it allows us to interpret the intercept a_0 as the payoff difference between Denmark and Norway at the average Norwegian payoff. The interpretation of the slope a_1 – which quantifies the average increase in the Danish payoffs for a one unit increase in the Norwegian payoffs – is unaffected by the centering.

There are two considerations that we need to pay attention to when taking equation (13) to the data: measurement error and across-population comparison. Unweighted estimation of (13) would assume that the estimated returns are from populations of similar size. In practice, the return estimates in the two countries will have differently sized groups, where some estimates are based on many applicants shifted by the instrument (when there are many applicants with given preferred and next-best fields and the first stage is large), while others are based on few applicants shifted (when there are less applicants in the preferred/next-best field cell or the first stage is close to zero).

To take these unequal underlying population sizes into account we will weigh our regressions with a measure of the number of applicants that are shifted. For each payoff estimate β_{jl}^{c} in country c we calculate the net number of applicants that are shifted on that margin as follows

$$n_{jl}^{k,c} = |\alpha_{jl}^{k,c}| \cdot N_{kl}^{c} \cdot \bar{z}_{kl}^{c}$$

where α_{jl}^{k} is the first-stage coefficient, N_{kl} the number of applicants with preferred field k and next-best field l, and $\bar{z}$ the share of these applicants above the cutoff. We then

[10] In both countries, the number of applicants with Medicine as next-best is very small and these are therefore omitted in our analysis. Thus, there are 9 preferred fields but only 8 next-best fields.

construct weights[11,12]

$$w_{jl} = \sum_k (n_{jl}^{k,NO} + n_{jl}^{k,DK})$$

Measurement error concerns arise because rather than relating population payoffs as in (13) we will be comparing two sets of noisily estimated population payoffs:

$$(\hat{\beta}_{jl}^{DK} - \bar{\hat{\beta}}^{NO}) = a_0 + a_1(\hat{\beta}_{jl}^{NO} - \bar{\hat{\beta}}^{NO}) + \tilde{e}_{jl} \tag{14}$$

It is well know that measurement error in explanatory variables results in estimation bias. Assuming classical measurement error $\hat{\beta}_{jl}^c = \beta_{jl}^c + \varepsilon_{jl}^c$ with ε_{jl}^c i.i.d. and $\sigma_{\varepsilon,c}^2 \equiv var(\varepsilon_{jl}^c)$, we can quantify the bias as follows[13]

$$\hat{a}_1 = \frac{cov(\hat{\beta}_{jl}^{DK}, \hat{\beta}_{jl}^{NO})}{var(\hat{\beta}_{jl}^{NO})} \to a_1 \frac{var(\beta_{jl}^{NO})}{var(\beta_{jl}^{NO}) + var(\varepsilon_{jl}^{NO})} = a_1 R_{NO}$$

where the estimate of a_1 is attenuated by a factor $R_{NO} = 1 - \sigma_{\varepsilon,NO}^2 / \sigma_{\hat{\beta},NO}^2$ (with $\sigma_{\hat{\beta},NO}^2 \equiv var(\hat{\beta}_{jl}^{NO})$). R_{NO} quantifies the reliability of $\hat{\beta}_{jl}^{NO}$ and, provided we can estimate it, implies that we can adjust $\hat{a}_1$ by $1/\hat{R}_{NO}$ to recover an unbiased estimate of the true a_1.[14] We construct an estimate of R_{NO} by plugging in the variance of the payoff estimates as an estimate of $\sigma_{\hat{\beta},c}^2$, and using the average squared standard errors of the payoffs as an estimate of $\sigma_{\varepsilon,c}^2$.[15] Finally, we can use the so-called total reliability $R_{Total} = \sqrt{R_{NO} \cdot R_{DK}}$ to construct an estimate of the correlation of the payoffs across the two countries

$$\hat{\rho} = \rho(\hat{\beta}_{jl}^{DK}, \hat{\beta}_{jl}^{NO}) / \hat{R}_{Total} \to \rho \equiv \rho(\beta_{jl}^{DK}, \beta_{jl}^{NO})$$

Table 3 reports the standard deviation of the estimated payoffs, the square root of their average standard errors squared, as well as the resulting estimated reliability ratios. The first two columns report the unweighted estimates. We see that the payoff estimates vary more in Norway than in Denmark and are on average also more noisily estimated. These unweighted estimates do however not map into a population. The analysis in this paper will therefore investigate weighted results and the next two columns report the weighted reliability estimates. For shifted applicants the variability in the estimates and the average standard error is reduced, especially for the Norwegian estimates. The estimated relia-

[11] It should be noted that in practice using these weights gives very similar results to using population weights $N_{kl}^{NO} + N_{kl}^{DK}$.

[12] When we study distributions of first-stage coefficients we will also use the weights w_{jl}.

[13] Classical measurement error in the dependent variable affects the precision but not the consistency of the regression estimates.

[14] We use the Stata command -eivreg- to perform the error-in-variable regression.

[15] Sullivan (2001) shows that this approach is robust to measurement error heteroskedasticity.

Table 3. Descriptive statistics and reliabilities, Norwegian and Danish payoff estimates

	Unweighted		Weighted	
	Norway	Denmark	Norway	Denmark
SD of payoff estimates $\hat{\beta}^c_{jl}$ ($\hat{\sigma}_{\hat{\beta},c}$)	40.4	20.0	31.6	18.6
Square root of average $SE(\hat{\beta}^c_{jl})^2$ ($\hat{\sigma}_{\varepsilon,c}$)	29.0	11.2	11.8	9.9
Reliability ($R_c = 1 - \hat{\sigma}^2_{\varepsilon,c}/\hat{\sigma}^2_{\hat{\beta},c}$)	0.48	0.71	0.86	0.72
Total reliability ($R_{Total} = \sqrt{R_{NO} \cdot R_{DK}}$)	0.58		0.79	

Note: See section 6.2 for the definition of the first-stage weights ω_{jl}.

bility of the Norwegian payoff estimates is 0.86 compared to 0.72 for the Danish ones. Reliability is therefore high for both countries.[16]

7 Payoffs to fields of study

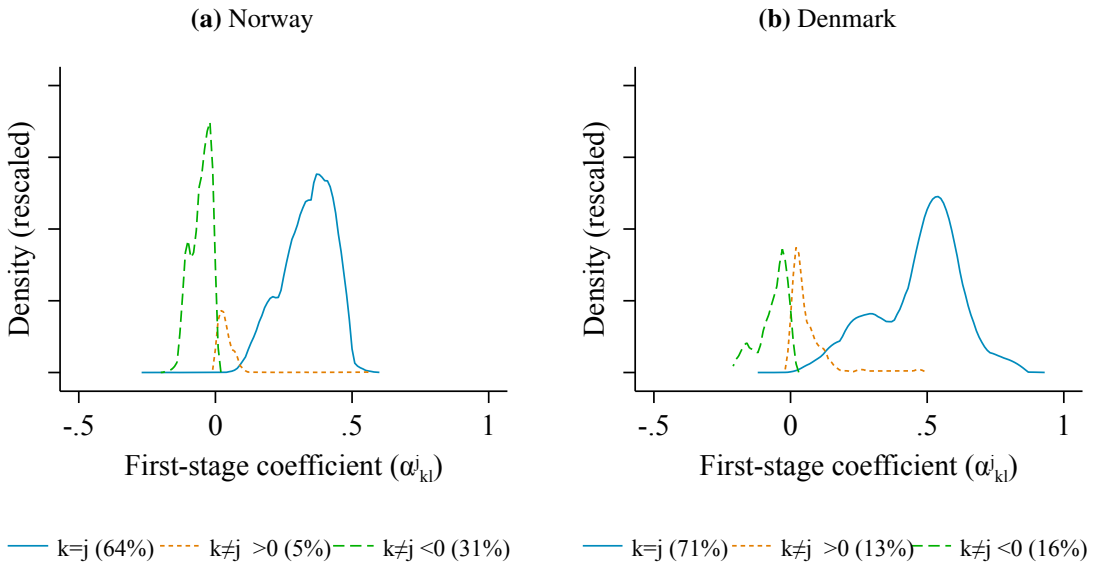
7.1 Examining the violation of next-best and irrelevance

We start by examining whether we can statistically reject the irrelevance and/or next best assumptions. As shown above, these assumptions are rejected if any of the off-diagonal first-stage coefficients are significantly different from zero. Joint tests strongly reject this null-hypothesis for both countries (cf. Tables A4 and A6). This implies that there are irrelevance-violators (detected by positive off-diagonal first-stage coefficients) and/or next-best-violators (detected by negative off-diagonal first-stage coefficients). We tend to detect such violation for most field of studies.

As a first indication of the relative importance of next-best vs. irrelevance violations we consider the signs of the off-diagonal coefficients that are individually significant (Tables A5 and A7). For Norway this reveals that few if any of the positive coefficients are individually significant, especially after adjusting for multiple testing (using the Bonferroni correction). However, a large number of the negative off-diagonal coefficients are significant, also after adjusting for multiple testing. For Norway, we therefore mostly find evidence for violations of next-best. This stands in contrast to the results for the Danish data, which are consistent with violations of the irrelevance and next-best assumptions being approximately equally frequent.

With enough data any model can be rejected, no matter how minor the misspecification. We therefore gauge the empirical relevance of the violations of the irrelevance and

[16] Using country-specific weights gives slightly higher but very similar estimates, namely a reliability of 0.89 for Norway and 0.78 for Denmark.

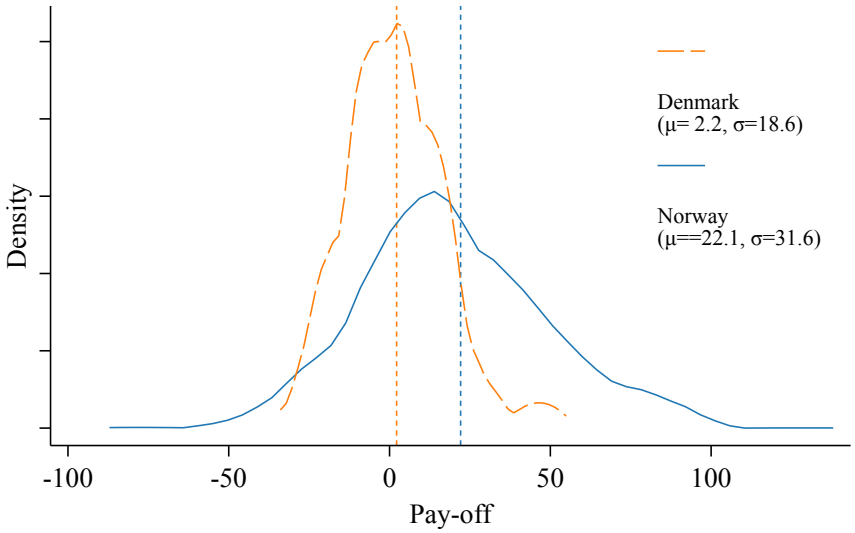


Note: Weighted densities that sum together to 1. First-stage coefficients α^k_{jl} where $k = j$ are "on-diagonal". Those with $l \neq j$ are "off-diagonal", and can be either positive (">0") or negative ("<0"). Percentage shares are indicated in parentheses.

Figure 5. Distribution of first-stage coefficients

next-best alternative conditions by quantifying the relative size of the associated applicant groups. The results are reported in Figure 5, which shows the distribution of the relevant first-stage coefficients weighted with the number of applicants shifted and where the densities are rescaled so that they to sum to unity. The mass under each density – reported in parenthesis in the Figure – quantifies the relative size of the complier/defier group in question.

The left-panel of Figure 5 shows that at least 64% of the shifted applicants in Norway are shifted at the expected (on-diagonal) margin. Of the remaining shifted applicants nearly 90% are shifted on margins with negative coefficients. For Norway we therefore continue to find evidence for violations of next-best but not irrelevance when we take the size of the shifted applicant groups into account. The results for Denmark in the right-panel of Figure 5 show that a similar share of applicants is shifted at the diagonal. Off-diagonal the shifted applicants are however evenly distributed between positive and negative margins. This reinforces the earlier conclusion, suggesting that violations of the irrelevance and next-best assumptions are approximately equally frequent. It should be emphasized however that, depending on their sign, the (absolute values of the) off-diagonal first-stage coefficients give a lower bound on each type of violator, while the on-diagonal coefficients provide upper bounds on the compliers. We therefore conclude that in both countries the violations of irrelevance or next-best are quantitatively non-trivial,



Note: Weighted distributions, compressed with a factor of R along the x-axis, and expanded by a factor 1/R along the y-axis (to maintain an area of one). Completed/next-best pairs (64) are the same in both countries. μ=Mean, σ=Std.dev.

Figure 6. Distribution of payoffs by country

appear to be of similar magnitude, but of a different nature.

7.2 Comparison of payoffs

Figure 6 reports the reliability-corrected and weighted densities for the Norwegian and Danish estimates of the payoffs of completing a field-of-study instead of the next-best.[17] In each country, the payoffs are measured in terms of annual earnings eight years after application. On average the annual payoff in Denmark is about 2,200 USD, while in Norway the returns are substantially larger at about 22,000 USD. In addition, there is also more variation in the payoffs in Norway compared to Denmark. A joint test of equality of the payoffs across countries gives a χ^2_{64} statistic of 441.7 with a corresponding p-value smaller than 0.0001. We therefore strongly reject that the payoffs are the same. In the following we investigate these differences in more detail.

Figure 7 starts out with comparing the Norwegian and Danish payoff estimates directly. It plots the estimates in the two countries against each other, with the size of the marker being proportional to the size of the sum of the Norwegian and Danish shifted applicant groups and, in addition to the 45-degree line, the figure also shows the regression line from the following error-in-variables regression (14) described in section 6.2 above:

$$(\hat{\beta}_{jl}^{DK} - \overline{\hat{\beta}}^{NO}) = a_0 + a_1(\hat{\beta}_{jl}^{NO} - \overline{\hat{\beta}}^{NO}) + \tilde{e}_{jl} \tag{15}$$

[17] Appendix tables A8 and A9 report the payoff estimates.

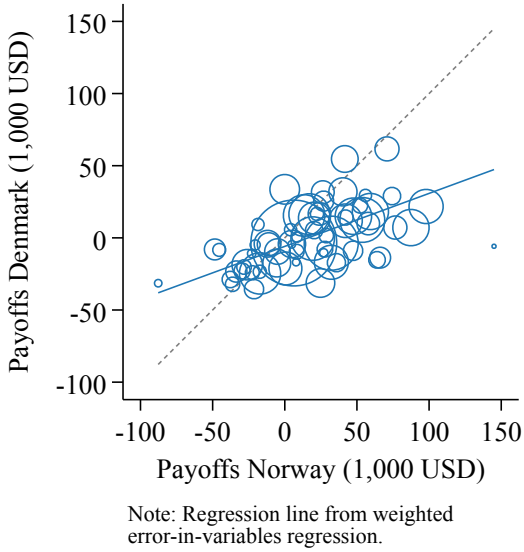


Figure 7. Payoffs in Denmark and Norway, all completed and next-best fields

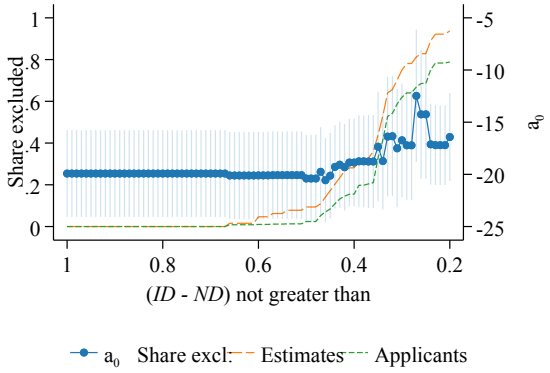
Changes in the intercept a_0 as we omit estimates with evidence of defiance shows whether the average Danish payoff become more aligned with the average Norwegian payoffs. We also report changes in the slope a_1 and the estimated correlation between the Danish and Norwegian payoffs ρ.

Figure 8 reports the results of this exercise and shows that, consistent with the low average and lower spread of the Danish estimates in Figure 6, the Danish estimates increase less than one-to-one with the Norwegian estimates with an estimated slope of 0.38 (s.e. 0.07), and are on average substantially lower (the estimated payoff difference is -19.9 with a s.e. of 2.1). However, even though their levels are different, we find that the payoffs exhibit a relatively strong positive correlation of 0.65 after adjusting for measurement error.

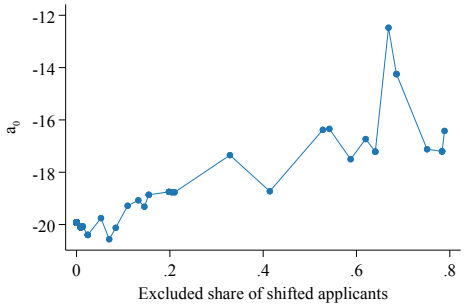
Above we found evidence of violations of the irrelevance and next-best assumptions for both two countries. Can these violations explain the observed differences in the estimated payoffs? We investigate this question by successively removing preferred-next-best combinations with a high share of detected defiers, thus reducing the share of defiers in the sample and see how this impacts the relationships between the Norwegian and Danish payoff estimates.

We first compute, for each completed and next-best field in both countries, the share of applicants that are shifted by off-diagonal instruments. This quantifies the net-flow of irrelevance and next-best defiers at that particular margin. We then progressively drop the estimates with the largest shares of net-defiance and estimate the weighted error-in-variables regression on the resulting sub-sample of Danish payoffs on Norwegian payoffs.

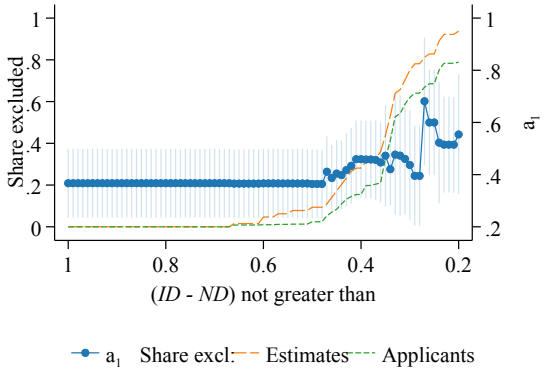
(a) Changes in a_0, # estimates and shifted applicants



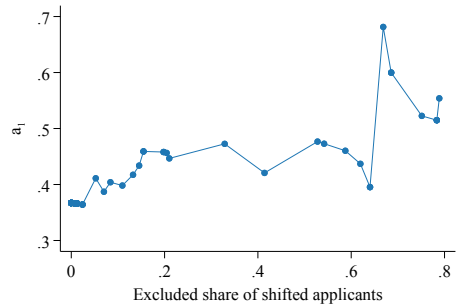
(b) a_0 vs. exclusion of violators



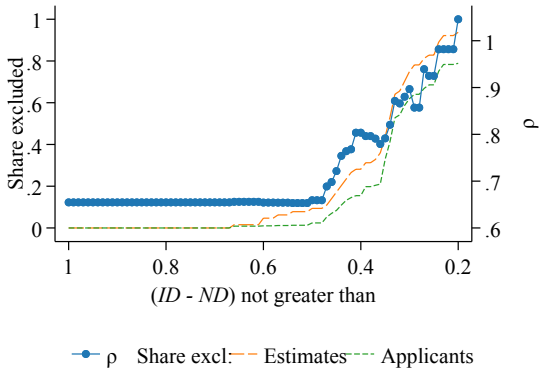
(c) Changes in a_1, # estimates and shifted applicants



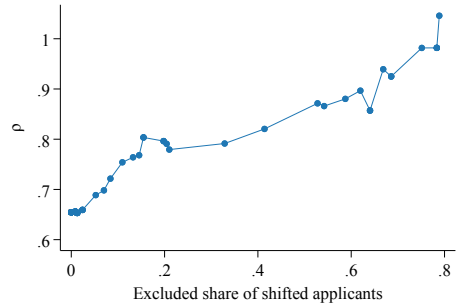
(d) a_1 vs. exclusion of violators



(e) Changes in payoff correlation ρ, # estimates and shifted applicants



(f) Correlation vs. exclusion of violators



Note: Graphs show (a, b) constant term from a weighted error-in-variable-regression of Danish payoffs on Norwegian payoffs, (c, d) coefficient on Norwegian payoffs from the same regression, and (e, f) the weighted reliability-adjusted correlation between Danish and Norwegian payoff estimates.

Figure 8. Regression coefficients and correlations as a function of defiers excluded

Sub-graph (a) shows the estimated intercepts and their confidence intervals. The x-axis in sub-graph (a) shows the maximum share of net-defiance allowed in the sample of estimates for the two countries. Reducing this share one percentage point at a time we re-estimate (15). The first estimate is dropped at about 66 percent net-defiance. Then progressively more payoff estimates are excluded as we restrict the maximum share of defiers below 50 percent. We see that a_0 stays approximately constant close to -20 until we restrict the share of defiers to be below 50 percent. After this a_0 gradually increases, and reaches -17 when we restrict the sample to max 20 percent defiers.

Sub-graph (a) also shows the shares of the 64 payoff estimates and of the shifted applicants that are excluded. The pairs of completed/next-best fields that have the highest shares of net-defiers have relatively few applicants shifted on the diagonal. Restricting the maximum to 50 percent we exclude 9 percent of estimates and 2 percent of compliers. Restricting further has a stronger impact on estimates and shifted applicants retained, and when we ultimately restrict the sample to max 20 percent defiers only 4 out of 64 estimates and 21 percent of the shifted applicants are retained.

In sub-graph (b) we plot a_0 against the share of shifted applicants that are excluded. As a function of applicants excluded, a_0 rises about linearly. However, as can be seen from the confidence bands in sub-graph (a), the estimated intercepts for different samples are never significantly different.

In sub-graphs (c) and (d) we show similar results for the slope parameter a_1 from (15). While a_1 increases somewhat in the beginning, it is mostly stable across the different samples. Finally, in sub-graphs (e) and (f) we show the reliability-adjusted weighted coefficient of correlation. This increases steadily with the share of compliers excluded, from 0.65 in the full sample to 1 when restricting to less than 20 percent net-defiers.

While we found above that the Norwegian and Danish payoff estimates are strongly correlated, this correlation substantially increases further when we exclude the estimates with more evidence of defiance of irrelevance and next-best. The intercept and slope from the regression (15) are however relatively stable, suggesting that violations of irrelevance and next-best do not explain the lower level and variation of the payoffs in Denmark compared to Norway.

7.3 Other explanations for differences in payoffs across the countries

To explore other explanations for the between-country differences in payoffs, we re-estimate (15) while adjusting for completed and next-best field dummies, as well as differences in average selectivity and earnings (cf. Figure 4) across completed and next-best fields. Table 4 reports the results. The first column reproduces the basic results reported above in Figure 7 where we found that the payoff difference was about 20,000 USD, and

Table 4. Explaining payoff differences between Denmark and Norway

	Earnings at $t=8$						Earnings at $t=13$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
a_0	-19.92	-19.92	-19.92	-15.62	-20.39	-15.70	-10.25	-4.31
	(2.12)	(1.56)	(1.71)	(3.15)	(2.01)	(3.15)	(3.26)	(5.52)
a_1	0.37	0.29	0.70	0.56	0.41	0.56	0.70	0.91
	(0.07)	(0.07)	(0.13)	(0.10)	(0.08)	(0.10)	(0.15)	(0.17)
Controls X_{jl}								
- Completed field		✓						
- Next-best field			✓					
- ΔGPA				✓		✓		✓
- ΔEarnings					✓	✓		
R^2	0.34	0.72	0.64	0.51	0.43	0.52	0.34	0.54

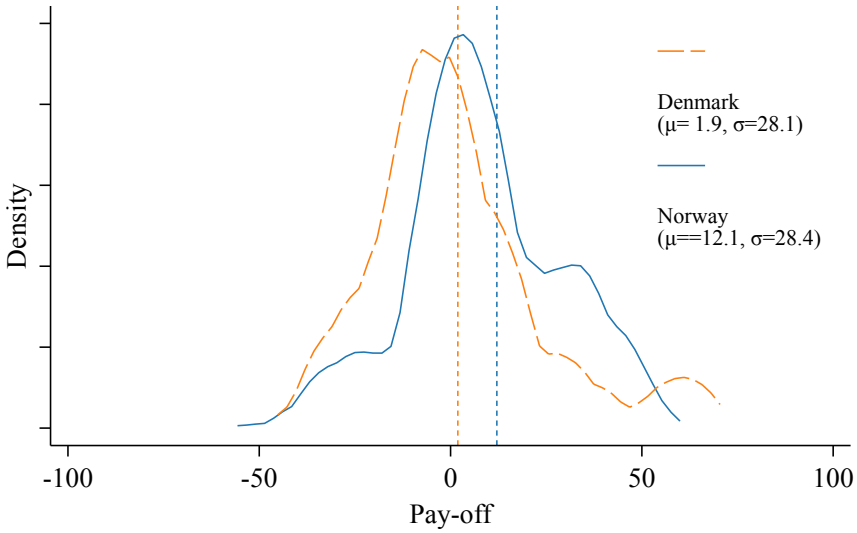
Note: Weighted error-in-variable estimates of $(\hat{\beta}_{jl}^{DK} - \overline{\hat{\beta}}^{NO}) = a_0 + a_1(\hat{\beta}_{jl}^{NO} - \overline{\hat{\beta}}^{NO}) + X_{jl}\gamma + \tilde{e}_{jl}$. $N = 64$ for $t = 8$, $N = 61$ for $t = 13$. Standard errors in parentheses.

that the payoffs in Denmark increased by less than one for each unit increase in Norway reflecting the smaller variance in the payoff distribution in Denmark.

We next investigate whether payoffs are more aligned across completed fields or across next-best fields. The next two columns of Table 4 therefore adjust for completed field and next-best field dummies. Keeping completed field fixed we now obtain a slope estimate of 0.29 in column (2), while keeping next-best field fixed in column (3) increases the slope substantially to 0.70. This shows that differences between next-best fields contribute more to between-country differences in payoffs than differences between completed fields.

In a next step we investigate two potential explanation of such differences. First we verify whether differential selectivity plays a role by adjusting for across country differences in average GPA in both the completed field j and the next-best field l. This changes the interpretation of the intercept which now corresponds to the across country payoff difference keeping the average GPA the same in the completed and next-best field. The estimates in column (4) show that this reduces the payoff gap with 25% from about -20,000 to -16,000 USD, while at the same time payoff become more evenly distributed as shown by the increase in the slope coefficient from 0.37 to 0.56. A similar exercise using average earnings in column (5) and (6) shows that this does not explain across country differences.

As noted earlier, looking at earnings eight years after application corresponds to rel-



Note: Weighted distributions, compressed with a factor of R along the x-axis, and expanded by a factor 1/R along the y-axis (to maintain an area of one). Completed/next-best pairs (61) are the same in both countries. μ=Mean, σ=Std.dev.

Figure 9. Distribution of longer-run payoffs by country (13 years since applying)

atively early career outcomes, especially for 5-year programs and studies that are not closely tied to a narrow set of occupations and which may therefore have longer and more complex school-to-work transitions. We therefore also compare the Danish payoff estimates 13 years after applying with the Norwegian estimates 13 years after applying. Extending the time horizon by 5 years has some impact on the estimated reliability of the Norwegian estimates which drops to 0.64, but the raw correlation between the $t = 8$ and $t = 13$ estimates is high (0.80).[18] For Denmark the reliability is slightly higher at $t = 13$ as is the raw correlation between the $t = 8$ and $t = 13$ estimates (0.86).

Figure 9 shows the estimated distributions of these longer-run payoffs across the two countries. Compared to the early career payoffs, the Norwegian and Danish payoff distribution are now much more aligned in terms of location and scale. This can also be seen in column (7) of Table 4. The average payoff gap between the two countries is now about 10,000 USD, and the slope coefficient has increased to 0.70.[19] Column (8) shows that after adjusting for differential selectivity the payoff estimates are on average aligned and we cannot reject that the intercept equals zero and the slope equals one (the corresponding F-test gives a p-value of 0.67). However, the estimated (reliability corrected) correlation coefficient between the Norwegian and Danish payoff estimates barely moves when comparing $t = 8$ vs. $t = 13$ (0.66 vs 0.65).

[18] We need to exclude three very imprecisely estimated payoffs with Law as next-best field from the $t = 13$ analysis (see appendix Table A10) to recover a non-negative reliability estimate.

[19] Appendix Tables A10 and A11 report the estimates, and appendix Figure A5 compares the payoff estimates and reports the error-in-variables regression line.

To summarize, we find that payoff estimates are strongly correlated across countries but have initially different levels and dispersion. Violations of the irrelevance and next-best assumptions that underpin the empirical approach do weaken the correlation, but appear to have little consequence for the estimated level and variance differences. Over time, the level and variance difference converge across countries, but this does not affect the correlation of the payoffs. Additional exploratory analyses show that these across country differences are mostly driven by heterogeneity in next-best fields which can partly be explained by differences in selectivity.

8 Conclusion

We revisited the identification argument of Kirkeboen et al. (2016) who showed how one may combine instruments for multiple unordered treatments with information about individuals' ranking of these treatments to achieve identification while allowing for both observed and unobserved heterogeneity in treatment effects. We showed that the key assumptions underlying their identification argument have testable implications. We also provided a new characterization of the bias that may arise if these assumptions are violated. Taken together, these results allow researchers not only to test the underlying assumptions, but also to argue whether the bias from violation of these assumptions are likely to be economically meaningful.

Guided and motivated by these results, we estimated and compared the earnings payoffs to post-secondary fields of study in Norway and Denmark. In each country, we applied and assessed the identification argument of Kirkeboen et al. (2016) to data on individuals' ranking of fields of study and field-specific instruments from discontinuities in the admission systems. We empirically examined whether and why the payoffs to fields of study differ across the two countries. We found strong cross-country correlation in the payoffs to fields of study, especially after removing fields with violations of the assumptions underlying the identification argument.

While our empirical findings are specific to the context of postsecondary education in the Nordic countries, there could be lessons from our work for other settings with unordered choices. Our study highlights key challenges and possible solutions to understanding what the causal effects of these choices are. Examples can be found in observational studies that use IV to study workers' selection of occupation, students' choice of education, firms' decision on location, or families' choice of where to live. Another example is the frequent use of IV to analyze encouragement designs in experiments where treatments are made available but take up is not universal (Duflo et al., 2007).

References

Altonji, J. G., Arcidiacono, P., and Maurel, A. (2016). The analysis of field choice in college and graduate school: Determinants and wage effects. In *Handbook of the Economics of Education*, volume 5, chapter 7, pages 305–396. Elsevier.

Altonji, J. G., Blom, E., and Meghir, C. (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics*, 4(1):185–223.

Duflo, E., Glennerster, R., and Kremer, M. (2007). Using randomization in development economics research: A toolkit. In *Handbook of development economics*, volume 4, chapter 61, pages 3895–3962. Elsevier.

Heckman, J. J. and Urzúa, S. S. (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics*, 156(1):27–37.

Heckman, J. J., Urzúa, S. S., and Vytlacil, E. J. (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics*, 88(3):389–432.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Kamat, V. (2017). Identification with latent choice sets: The case of the head start impact study. *arXiv preprint arXiv:1711.02048*, 1.

Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111.

Lee, S. and Salanié, B. (2020). Filtered and unfiltered treatment effects with targeting instruments. *arXiv preprint arXiv:2007.10432*.

Mogstad, M. and Torgovitsky, A. (2018). Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics*, 10:577–613.

Nibbering, D., Oosterveen, M., and Silva, P. L. (2022). Clustered local average treatment effects: fields of study and academic student progress. Discussion Paper No. 15159, IZA Institute for Labor Economics.

Sullivan, D. G. (2001). A note on the estimation of linear regression models with heteroskedastic measurement errors. Working paper 2001-23, Federal Reserve Board of Chicago.

Svensson, L.-G. (1999). Strategy-proof allocation of indivisible goods. *Social Choice and Welfare*, 16(4):557–567.

Appendix

A Proof Of Bias When Auxiliary Assumptions Fail

Proof. We build on the notation from Section 2.1. IV uses the three moment conditions:

$$\mathbb{E}[\varepsilon] = 0, \quad \mathbb{E}[\varepsilon z_1] = 0 \quad \text{and} \quad \mathbb{E}[\varepsilon z_2] = 0$$

Expressing ε in terms of potential outcomes, we get:

$$\begin{aligned} \varepsilon &= (y^0 - \beta_0) + (y^1 - y^0 - \beta_1)d_1 + (y^2 - y^0 - \beta_2)d_2 \\ &= (y^0 - \beta_0) + (y^1 - y^0 - \beta_1)(d_1^0 + (d_1^1 - d_1^0)z_1 + (d_1^2 - d_1^0)z_2) \\ &\quad + (y^2 - y^0 - \beta_2)(d_2^0 + (d_2^1 - d_2^0)z_1 + (d_2^2 - d_2^0)z_2) \end{aligned} \tag{16}$$

We substitute into the moment conditions, and solve. Under independence, we get:

$$\begin{aligned} \mathbb{E}[(y^1 - y^0 - \beta_1)(d_1^1 - d_1^0) + (y^2 - y^0 - \beta_2)(d_2^1 - d_2^0)] = 0 \\ \mathbb{E}[(y^1 - y^0 - \beta_1)(d_1^2 - d_1^0) + (y^2 - y^0 - \beta_2)(d_2^2 - d_2^0)] = 0 \end{aligned}$$

As shown by Kirkeboen et al. (2016), this implies, for $k = 1,2$, $k' = 2,1$, that:

$$\begin{aligned} &\mathbb{E}[y^k - y^0 - \beta_k \mid d_k^k - d_k^0 = 1, d_{k'}^k - d_{k'}^0 = 0] \times P[d_k^k - d_k^0 = 1, d_{k'}^k - d_{k'}^0 = 0] \qquad (17) \\ &+ \mathbb{E}[(y^k - y^0 - y^{k'} - y^0) - (\beta_k - \beta_{k'}) \mid d_k^k - d_k^0 = 1, d_{k'}^k - d_{k'}^0 = -1] \times P[d_k^k - d_k^0 = 1, d_{k'}^k - d_{k'}^0 = -1] \\ &+ \mathbb{E}[y^{k'} - y^0 - \beta_{k'} \mid d_k^k - d_k^0 = 0, d_{k'}^k - d_{k'}^0 = 1] \times P[d_k^k - d_k^0 = 0, d_{k'}^k - d_{k'}^0 = 1] = 0 \qquad (18) \end{aligned}$$

where we have assumed

$$P[d_k^k - d_k^0 = -1, d_{k'}^k - d_{k'}^0 = 0] = P[d_k^k - d_k^0 = 0, d_{k'}^k - d_{k'}^0 = -1] = 0$$

under monotonicity. To simplify notation, we rewrite equation 17 in terms of the notation from Table 1:

$$\begin{aligned} &\mathbb{E}[y^k - y^0 - \beta_k \mid C_k] \times P(C_k) \\ &+ \mathbb{E}[(y^k - y^0 - y^{k'} - y^0) - (\beta_k - \beta_{k'}) \mid ND_k] \times P(ND_k) \\ &+ \mathbb{E}[y^{k'} - y^0 - \beta_{k'} \mid ID_k] \times P(ID_k) = 0 \end{aligned}$$

We isolate β_k for $k = 1, 2$:

$$\beta_k = \beta_{k'} \frac{P(ND_k) - P(ID_k)}{P(C_k) + P(ND_k)} + \frac{\mathbb{E}[y^k - y^0 \mid C_k] P(C_k)}{P(C_k) + P(ND_k)} \tag{19}$$

$$+ \frac{\mathbb{E}[y^k - y^0 - y^{k'} - y^0 \mid ND_k] P(ND_k)}{P(C_k) + P(ND_k)} + \frac{\mathbb{E}[y^{k'} - y^0 \mid ID_k] P(ID_k)}{P(C_k) + P(ND_k)}$$

□

A.1 No Auxiliary Assumptions

We substitute equation (19) with $k = 2$ into (19) with $k = 1$ and get:

$$\beta_1 = \frac{\mathbb{E}[y^1 - y^0 \mid C_1] P(C_1)}{P(C_1) + P(ND_1)}$$

$$+ \mathbb{E}[y^2 - y^0 \mid ID_1] \frac{P(ID_1)}{P(C_1) + P(ND_1)} + \frac{\mathbb{E}[y^1 - y^0 - y^2 - y^0 \mid ND_1] P(ND_1)}{P(C_1) + P(ND_1)}$$

$$+ \frac{P(ND_1) - P(ID_1)}{P(C_1) + P(ND_1)} \times \left[\frac{\mathbb{E}[y^2 - y^0 \mid C_2] P(C_2)}{P(C_2) + P(ND_2)} + \mathbb{E}[y^1 - y^0 \mid ID_2] \frac{P(ID_2)}{P(C_2) + P(ND_2)} \right.$$

$$\left. + \frac{\mathbb{E}[y^2 - y^0 - y^1 - y^0 \mid ND_2] P(ND_2)}{P(C_2) + P(ND_2)} + \beta_1 \frac{P(ND_2) - P(ID_2)}{P(C_2) + P(ND_2)} \right]$$

Letting

$$\dot{W} = 1 - \frac{(P(ND_1) - P(ID_1))(P(ND_2) - P(ID_2))}{(P(C_1) + P(ND_1))(P(C_2) + P(ND_2))}$$

$$= \frac{(P(C_1) + P(ND_1))(P(C_2) + P(ND_2)) - (P(ND_1) - P(ID_1))(P(ND_2) - P(ID_2))}{(P(C_1) + P(ND_1))(P(C_2) + P(ND_2))}$$

and gathering β_1-terms on the LHS gives:

$$\beta_1 \dot{W} = \mathbb{E}[y^1 - y^0 \mid C_1] \times \frac{P(C_1)}{P(C_1) + P(ND_1)}$$

$$+ \mathbb{E}[y^2 - y^0 \mid ID_1] \times \frac{P(ID_1)}{P(C_1) + P(ND_1)}$$

$$+ \mathbb{E}[y^1 - y^0 - y^2 - y^0 \mid ND_1] \times \frac{P(ND_1)}{P(C_1) + P(ND_1)}$$

$$+\mathbb{E}[y^1-y^0 \mid ID_2]\times\frac{(P(ND_1)-P(ID_1))P(ID_2)}{(P(C_1)+P(ND_1)(P(C_2)+P(ND_2))}$$

$$+\mathbb{E}[y^2-y^0 \mid C_2]\times\frac{(P(ND_1)-P(ID_1))P(C_2)}{(P(C_1)+P(ND_1))(P(C_2)+P(ND_2))}$$

$$+\mathbb{E}[y^2-y^0-y^1-y^0 \mid ND_2]\times\frac{(P(ND_1)-P(ID_1))P(ND_2)}{(P(C_1)+P(ND_1))(P(C_2)+P(ND_2))}$$

Adding and subtracting

$$\mathbb{E}[y^1-y^0 \mid C_1]\frac{P(ND_1)}{P(C_1)+P(ND_1)}+\mathbb{E}[y^1-y^0 \mid C_1]\frac{(P(ND_1)-P(ID_1))(P(ND_2)-P(ID_2))}{(P(C_1)+P(ND_1))(P(C_2)+P(ND_2))}$$

on the RHS and gathering terms gives:

$$\begin{aligned}\beta_1\dot{W}=\mathbb{E}[y^1-y^0 \mid C_1]\dot{W} \quad - \quad & \mathbb{E}[y^1-y^0 \mid C_1]\times\frac{P(ND_1)(P(C_2)+P(ND_2))}{(P(C_1)+P(ND_1))(P(C_2)+P(ND_2))}\\ & +\mathbb{E}[y^1-y^0 \mid C_1]\times\frac{(P(ND_1)-P(ID_1))(P(ND_2)-P(ID_2))}{(P(C_1)+P(ND_1))(P(C_2)+P(ND_2))}\\ & +\mathbb{E}[y^2-y^0 \mid ID_1]\times\frac{P(ID_1)(P(C_2)+P(ND_2))}{(P(C_1)+P(ND_1))(P(C_2)+P(ND_2))}\\ & +\mathbb{E}[y^1-y^0-y^2-y^0 \mid ND_1]\times\frac{P(ND_1)(P(C_2)+P(ND_2))}{(P(C_1)+P(ND_1))(P(C_2)+P(ND_2))}\\ & +\mathbb{E}[y^1-y^0 \mid ID_2]\times\frac{(P(ND_1)-P(ID_1))P(ID_2)}{(P(C_1)+P(ND_1)(P(C_2)+P(ND_2))}\\ & +\mathbb{E}[y^2-y^0 \mid C_2]\times\frac{(P(ND_1)-P(ID_1))P(C_2)}{(P(C_1)+P(ND_1))(P(C_2)+P(ND_2))}\\ & +\mathbb{E}[y^2-y^0-y^1-y^0 \mid ND_2]\times\frac{(P(ND_1)-P(ID_1))P(ND_2)}{(P(C_1)+P(ND_1))(P(C_2)+P(ND_2))}\end{aligned}$$

Dividing by $\dot{W}$ on both sides, and letting

$$\bar{W}=(P(C_1)+P(ND_1))(P(C_2)+P(ND_2))-(P(ND_1)-P(ID_1))(P(ND_2)-P(ID_2))$$

gives

$$\beta_1=\mathbb{E}[y^1-y^0 \mid C_1] \quad - \quad \mathbb{E}[y^1-y^0 \mid C_1]\times\frac{P(ND_1)(P(C_2)+P(ND_2))}{\bar{W}}$$

$$
\begin{aligned}
&+\mathbb{E}[y^1-y^0 \mid C_1]\times\frac{(P(ND_1)-P(ID_1))(P(ND_2)-P(ID_2))}{\bar{W}}\\
&+\mathbb{E}[y^2-y^0 \mid ID_1]\times\frac{P(ID_1)(P(C_2)+P(ND_2))}{\bar{W}}\\
&+\mathbb{E}[y^1-y^0-y^2-y^0 \mid ND_1]\times\frac{P(ND_1)(P(C_2)+P(ND_2))}{\bar{W}}\\
&+\mathbb{E}[y^1-y^0 \mid ID_2]\times\frac{(P(ND_1)-P(ID_1))P(ID_2)}{\bar{W}}\\
&+\mathbb{E}[y^2-y^0 \mid C_2]\times\frac{(P(ND_1)-P(ID_1))P(C_2)}{\bar{W}}\\
&+\mathbb{E}[y^2-y^0-y^1-y^0 \mid ND_2]\times\frac{(P(ND_1)-P(ID_1))P(ND_2)}{\bar{W}}
\end{aligned}
$$

Rearranging, we get:

$$
\begin{aligned}
\beta_1^{IV} = \mathbb{E}[y^1-y^0 \mid C_1] \quad &+\frac{P(ND_1)P(C_2)}{\bar{W}} \times(\mathbb{E}[y^1-y^0 \mid ND_1]-\mathbb{E}[y^1-y^0 \mid C_1])\\
&+\frac{P(ND_1)P(C_2)}{\bar{W}} \times(\mathbb{E}[y^2-y^0 \mid C_2]-\mathbb{E}[y^2-y^0 \mid ND_1])\\
&+\frac{P(ND_1)P(ND_2)}{\bar{W}}\times(\mathbb{E}[y^1-y^0 \mid ND_1]-\mathbb{E}[y^1-y^0 \mid ND_2])\\
&+\frac{P(ND_1)P(ND_2)}{\bar{W}}\times(\mathbb{E}[y^2-y^0 \mid ND_2]-\mathbb{E}[y^2-y^0 \mid ND_1])\\
&+\frac{P(ID_1)P(ID_2)}{\bar{W}} \times(\mathbb{E}[y^1-y^0 \mid C_1]-\mathbb{E}[y^1-y^0 \mid ID_2])\\
&+\frac{P(ID_1)P(C_2)}{\bar{W}} \times(\mathbb{E}[y^2-y^0 \mid ID_1]-\mathbb{E}[y^2-y^0 \mid C_2])\\
&+\frac{P(ID_1)P(ND_2)}{\bar{W}}\times(\mathbb{E}[y^1-y^0 \mid ND_2]-\mathbb{E}[y^1-y^0 \mid C_1])\\
&+\frac{P(ID_1)P(ND_2)}{\bar{W}}\times(\mathbb{E}[y^2-y^0 \mid ID_1]-\mathbb{E}[y^2-y^0 \mid ND_2])\\
&+\frac{P(ND_1)P(ID_2)}{\bar{W}}\times(\mathbb{E}[y^1-y^0 \mid ID_2]-\mathbb{E}[y^1-y^0 \mid C_1])
\end{aligned}
\tag{20}
$$

where we can rearrange the denominator such that

$$
\begin{aligned}
\bar{W} = \;& P(C_1)P(C_2)+P(C_1)P(ND_2)+P(ND_1)P(C_2)\\
&+P(ND_1)P(ID_2)+P(ID_1)P(ND_2)-P(ID_1)P(ID_2)
\end{aligned}
$$

and the expression for β_2^{IV} follows by symmetry.

A.2 Assuming Only Next-best

We now want to find an expression of the bias assuming only next-best.

Proof. Next-best ensures $P(ND_1) = P(ND_2) = 0$. Equation 20 then reduces to

$$\begin{aligned}\beta_1^{IV} = \mathbb{E}[y^1 - y^0 \mid C_1] \quad &+ \frac{P(ID_1)P(ID_2)}{W'} \times (\mathbb{E}[y^1 - y^0 \mid C_1] - \mathbb{E}[y^1 - y^0 \mid ID_2]) \\ &+ \frac{P(ID_1)P(C_2)}{W'} \times (\mathbb{E}[y^2 - y^0 \mid ID_1] - \mathbb{E}[y^2 - y^0 \mid C_2])\end{aligned} \tag{21}$$

where

$$W' = P(C_1)(P(C_2) - P(ID_1)P(ID_2)$$

□

A.3 Assuming Only Irrelevance

We now want to find an expression of the bias assuming only irrelevance.

Proof. Irrelevance ensures $P(ID_1) = P(ID_2) = 0$. Equation 20 then reduces to

$$\begin{aligned}\beta_1^{IV} = \mathbb{E}[y^1 - y^0 \mid C_1] \quad &+ \frac{P(ND_1)P(C_2)}{\hat{W}} \times (\mathbb{E}[y^1 - y^0 \mid ND_1] - \mathbb{E}[y^1 - y^0 \mid C_1]) \\ &+ \frac{P(ND_1)P(C_2)}{\hat{W}} \times (\mathbb{E}[y^2 - y^0 \mid C_2] - \mathbb{E}[y^2 - y^0 \mid ND_1]) \\ &+ \frac{P(ND_1)P(ND_2)}{\hat{W}} \times (\mathbb{E}[y^1 - y^0 \mid ND_1] - \mathbb{E}[y^1 - y^0 \mid ND_2]) \\ &+ \frac{P(ND_1)P(ND_2)}{\hat{W}} \times (\mathbb{E}[y^2 - y^0 \mid ND_2] - \mathbb{E}[y^2 - y^0 \mid ND_1])\end{aligned} \tag{22}$$

where

$$\begin{aligned}\hat{W} &= (P(C_1) + P(ND_1))(P(C_2) + P(ND_2)) - P(ND_1)P(ND_2)) \\ &= P(C_1)P(C_2) + P(C_1)P(ND_2) + P(ND_1)P(C_2)\end{aligned}$$

□

Table A1. Detailed taxonomy of behavioral groups.

Potential Field Choice			Behavioral type	
d^0	d^1	d^2	z_1-stratum	z_2-stratum
0	1	2	C_1	C_2
0	1	1	C_1	ID_2
0	1	0	C_1	NT_2
0	0	0	NT_1	NT_2
0	0	2	NT_1	C_2
2	2	2	OT_1	AT_2
1	1	1	AT_1	OT_2
1	1	2	AT_1	ND_2
2	1	2	ND_1	AT_2
0	2	2	ID_1	C_2

Note: The table decomposes the behavioral groups from Table 1 into subgroups (strata) where the z_1 stratum is the group defined by their potential field choices when changing the instrument from 0 to 1, and the z_2 stratum is correspondingly defined for an instrument change from 0 to 2. The table shows the possible behavioral responses under all states of the instrument. Note that other takers OT_1 (OT_2) refers to global always takers of field 2 (1).

B Proof of Testable Implications

B.1 First Stage Quantities

We start by proving Proposition 2

Proof. We start by introducing a richer decomposition of behavioral groups, building on Table 1. This is presented in Table A1.

Focusing on $k = 1$, we take expectations on both sides in equation (7). As $\mathbb{E}[\nu_1] = 0$, we get:

$$\mathbb{E}[d_1] = \alpha_1^0 + \alpha_1^1 \times \mathbb{E}[z_1] + \alpha_1^2 \times \mathbb{E}[z_2] \tag{23}$$

We decompose the LHS into potential outcomes, using that $z_0 = 1 - z_1 - z_2$. Under independence we have:

$$\mathbb{E}[d_1] = \mathbb{E}[d_1^0] + \mathbb{E}[d_1^1 - d_1^0] \times \mathbb{E}[z_1] + \mathbb{E}[d_1^2 - d_1^0] \times \mathbb{E}[z_2] \tag{24}$$

Using Table 1, as groups are disjoint, we have

$$\begin{aligned}
\mathbb{E}[d_1^0] &= P(d_1^0 = 1) = P(AT_1) \\
\mathbb{E}[d_1^1 - d_1^0] &= P(d_1^1 - d_1^0 = 1) = P(C_1) + P(ND_1) \\
\mathbb{E}[d_1^2 - d_1^0] &= P(d_1^2 - d_1^0 = 1) - P(d_1^2 - d_1^0 = -1) = P(ID_2) - P(ND_2)
\end{aligned}$$

where we in both instances have assumed monotonicity and AT_1 denotes always takers. This turns equation (23) into:

$$\begin{aligned} & \alpha_1^0 - P(AT_1) \\ + \; & [\alpha_1^1 - (P(C_1) + P(ND_1))] \times \mathbb{E}[z_1] \\ + \; & [\alpha_1^2 - (P(ID_2) - P(ND_2))] \times \mathbb{E}[z_2] = 0 \end{aligned}$$

By the rank condition (and symmetry for $k = 2$), this implies:

$$P(AT_1) = \alpha_1^0 \qquad P(AT_2) = \alpha_2^0 \tag{25}$$

$$P(C_1) + P(ND_1) = \alpha_1^1 \qquad P(C_2) + P(ND_2) = \alpha_2^2 \tag{26}$$

$$P(ID_1) - P(ND_1) = \alpha_2^1 \qquad P(ID_2) - P(ND_2) = \alpha_1^2 \tag{27}$$

Since groups are disjoint we have

$$P(C_1) + P(AT_1) + P(NT_1) + P(OT_1) + P(ID_1) + P(ND_1) = 1 \tag{28}$$

$$P(C_2) + P(AT_2) + P(NT_2) + P(OT_2) + P(ID_2) + P(ND_2) = 1 \tag{29}$$

By combining equation (28) with equations (25)-(27) we get[20]

$$P(NT_1) = 1 - \alpha_1^0 - \alpha_2^0 - \alpha_1^1 - \alpha_2^1 \tag{30}$$

$$P(NT_2) = 1 - \alpha_1^0 - \alpha_2^0 - \alpha_2^2 - \alpha_1^2 \tag{31}$$

□

B.2 Partial Identification Of Defiers

We continue by proving Proposition 4

Proof. From Proposition 2, we get the following information on $P(ND_1)$:

$$P(ND_1) = \begin{cases} -\alpha_2^1 + P(ID_1) \\ \alpha_2^0 - P(OT_1) \\ \alpha_1^1 - P(C_1) \end{cases} \tag{32}$$

where the first line follows from equation (27), the second from (26) and the third from combining equation 28 with 30 and 25. From equation (27) we know that $P(ID_1) =$

[20] Where we use the following $AT_2 = OT_1 \cup ND_1$ and $AT_1 = OT_2 \cup ND_2$.

$\alpha_2^1 + P(ND_1)$. Combining this with the information in equation (32) we have:

$$P(ID_1) = \begin{cases} \alpha_2^1 + P(ND_1) \\ \alpha_2^1 + \alpha_2^0 - P(OT_1) \\ \alpha_2^1 + \alpha_1^1 - P(C_1) \end{cases}$$

This gives the following bounds on $P(ID_1)$ and $P(ND_1)$

$$\begin{aligned} P(ND_1) &\geq -\alpha_2^1 & \qquad P(ID_1) &\geq \alpha_2^1 \\ P(ND_1) &\leq \alpha_2^0 & \qquad P(ID_1) &\leq \alpha_2^1 + \alpha_2^0 \\ P(ND_1) &\leq \alpha_1^1 & \qquad P(ID_1) &\leq \alpha_2^1 + \alpha_1^1 \end{aligned}$$

where also, trivially, $P(ID_1), P(ND_1) \geq 0$. It follows that the bounds on $P(ID_1)$ are:

$$\begin{aligned} \max\{0, -\alpha_2^1\} \leq P(ND_1) &\leq \min\{\alpha_1^1, \alpha_2^0\} \\ \max\{0, \alpha_2^1\} \leq P(ID_1) &\leq \max\{0, \alpha_2^1 + \min\{\alpha_1^1, \alpha_2^0\}\} \end{aligned}$$

and results for instrument 2 are symmetric. □

B.3 Assuming Next-best

We now prove Corollary 1.

Proof. Assuming next-best, we have $P(ND_1) = P(ND_2) = 0$. This turns equation (26) into:

$$\begin{aligned} P(AT_1) &= \alpha_1^0 & \qquad P(AT_2) &= \alpha_2^0 \\ P(C_1) &= \alpha_1^1 & \qquad P(C_2) &= \alpha_2^2 \\ P(ID_1) &= \alpha_2^1 & \qquad P(ID_2) &= \alpha_1^2 \end{aligned}$$

□

B.4 Assuming Irrelevance

Lastly, we prove Corollary 2

Proof. Assuming irrelevance, we have $P(ID_1) = P(ID_2) = 0$. This turns equation (26) into:

$$P(AT_1) = \alpha_1^0 \qquad P(AT_2) = \alpha_2^0$$

$$P(C_1) = \alpha_1^1 + \alpha_2^1 \qquad P(C_2) = \alpha_2^2 + \alpha_1^2$$
$$P(ND_1) = -\alpha_2^1 \qquad P(ND_2) = -\alpha_1^2$$

□

C Proof Of Violation of Exclusion Under Clustering

In the following, we derive an expression for the IV estimand under binary clustering, as presented in Section 4.2.1.

C.1 Introduction

As mentioned in Section 4.2.1, we have the binary IV estimand in our set-up as:

$$\tilde{\beta}_1^{\text{IV}} = \frac{\theta_1}{\pi_1}$$

where θ_1 is the reduced form and π_1 is the first stage between when clustering treatments in two clusters, S_0 and S_1, and seeking to estimate the effect of going from the former to the latter. In the following we will derive a general expression for this estimand.

C.1.1 First Stage We have the first stage given by the relation

$$\tilde{d} = \pi_0 + \pi_1 \tilde{z} + \nu$$

Taking expectations on both sides with $\mathbb{E}[\nu] = 0$, we get

$$\mathbb{E}[\tilde{d}] = \pi_0 + \pi_1 \times \mathbb{E}[\tilde{z}_1]$$

Decomposing the LHS into potential outcomes using $\tilde{d} = \tilde{d}^0 + (\tilde{d}^1 - \tilde{d}^0) \times \tilde{z}$ we get:

$$\mathbb{E}[\tilde{d}] = \mathbb{E}[\tilde{d}^0] + \mathbb{E}[\tilde{d}^1 - \tilde{d}^0] \times \mathbb{E}[\tilde{z}] \tag{33}$$

i.e. we have

$$\pi_0 + \pi_1 \times \mathbb{E}[\tilde{z}] = \mathbb{E}[\tilde{d}^0] + \mathbb{E}[\tilde{d}^1 - \tilde{d}^0] \times \mathbb{E}[\tilde{z}] \tag{34}$$

C.1.2 Reduced Form With respect to the reduced form, we have:

$$\theta_1 = \mathbb{E}[y \mid \tilde{z} = 1] - \mathbb{E}[y \mid \tilde{z} = 0]$$

We substitute for potential outcomes with $y = \tilde{y}^0 \times (1 - \tilde{d}) + \tilde{y}^1 \times \tilde{d}$

$$\theta_1 = \mathbb{E}[\tilde{y}^0(1 - \tilde{d}) + \tilde{y}^1 \tilde{d} \mid \tilde{z} = 1] - \mathbb{E}[\tilde{y}^0(1 - \tilde{d}) + \tilde{y}^1 \tilde{d} \mid \tilde{z} = 0]$$

Since we do not assume cluster-level exclusion, we need to keep potential treatments and outcomes instrument-dependent. Rearranging we get:

$$\begin{aligned}\theta_1 = &\mathbb{E}[\tilde{y}^{0,1}\tilde{d}_0^1 \mid \tilde{z}=1] + \mathbb{E}[\tilde{y}^{1,1}\tilde{d}_1^1 \mid \tilde{z}=1] \\ &- \mathbb{E}[\tilde{y}^{0,0}\tilde{d}_0^0 \mid \tilde{z}=0] - \mathbb{E}[\tilde{y}^{1,0}\tilde{d}_1^0 \mid \tilde{z}=0]\end{aligned}$$

Rearranging, this becomes:

$$\begin{aligned}\theta_1 =&\mathbb{E}[\tilde{y}^{0,1} \mid \tilde{d}^1=0]P(\tilde{d}^1=0) + \mathbb{E}[\tilde{y}^{1,1} \mid \tilde{d}^1=1]P(\tilde{d}^1=1) \\ &- \mathbb{E}[\tilde{y}^{0,0} \mid \tilde{d}^0=0]P(\tilde{d}^0=0) - \mathbb{E}[\tilde{y}^{1,0} \mid \tilde{d}^0=1]P(\tilde{d}^0=1) \qquad (35)\end{aligned}$$

Under control clustering, we will have

$$S_1=\{1\},\ \ S_0=\{0,2\} \qquad \text{or} \qquad S_1=\{2\},\ \ S_0=\{0,1\}$$

and under treatment clustering we will have

$$S_1=\{1,2\},\ \ S_0=\{0\}$$

We will treat these scenarios separately, but focussing on the former control clustering scenario as these are symmetric.

C.2 Control Clustering

We have $S_1=\{1\},\ \ S_0=\{0,2\}$ and seek to find an expression of the first stage, reduced form and IV estimand. For brevity of notation, we use the taxonomy in Table A2 to denote complier and defier groups.

C.2.1 First Stage Applying the taxonomy to the expectation in equation (33), under field level monotonicity we get:

$$\mathbb{E}[\tilde{d}^1-\tilde{d}^0] = P[\tilde{d}^1-\tilde{d}^0=1] - P[\tilde{d}^1-\tilde{d}^0=-1] = P(\overline{C})$$

From equation (34) we hence have by the rank condition

$$\pi_{1,0} = P(\overline{C})$$

Table A2. Taxonomy of response groups under control clustering

Type	Cluster Level		Field Level			Group	
	$\tilde{d}^0$	$\tilde{d}^1$	d^0	d^2	d^1	Field	Cluster
Compliers	0	1	0		1	$\overline{C}$	C_1
	0	1		2	1		C_2
	0	1	2		1		ND_1
	0	1		0	1		ND_2
Never Takers	0	0	2		0	$\overline{NT}$	ID_1
	0	0		0	2		ID_2

Note: The table shows potential treatments for field and cluster instruments for groups impacted by the cluster instrument under control clustering. At the field level, d^0 indicates which treatment is taken given $Z = 0$,d^2 indicates which treatment is taken given $Z = 2$ and d^1 indicates which treatment is taken when $Z = 1$. The notation is equivalent at the cluster level. Relative to the clustered instrument, $\overline{C}$ are compliers and $\overline{NT}$ are never takers. Relative to the field instrument, C are compliers, ND are next-best defiers and ID are irrelevance defiers, all relative to some field level instrument corresponding to a treatment in S_1.

C.2.2 Reduced Form We use Table A2 to decompose the expectations in equation (35). Under independence and field level monotonicity, we get:

$$\begin{aligned}\theta_1 &= \mathbb{E}[\tilde{y}^{0,1} \mid \overline{NT}] \times P(\overline{NT}) \\ &+ \mathbb{E}[\tilde{y}^{1,1} \mid \overline{C}] \times P(\overline{C}) \\ &- \mathbb{E}[\tilde{y}^{0,0} \mid \overline{C} \cup \overline{NT}] \times P(\overline{C} \cup \overline{NT})\end{aligned}$$

Since sets are disjoint, we can rearrange:

$$\begin{aligned}\theta_1 &= \mathbb{E}[\tilde{y}^{1,1} - \tilde{y}^{0,0} \mid \overline{C}] \times P(\overline{C}) \\ &- \mathbb{E}[\tilde{y}^{0,1} - \tilde{y}^{0,0} \mid \overline{NT}] \times P(\overline{NT})\end{aligned}$$

Using Table A2 to turn cluster level groups into field level groups, changing outcome indices to reflect instruments relevant to the group in question, we get:

$$\begin{aligned}\theta_1 &= \mathbb{E}[y^{1,1} - y^{0,0} \mid C_1] \times P(C_1) \\ &+ \mathbb{E}[y^{1,1} - y^{2,2} \mid C_2] \times P(C_2) \\ &+ \mathbb{E}[y^{1,1} - y^{2,0} \mid ND_1] \times P(ND_1) \\ &+ \mathbb{E}[y^{1,1} - y^{0,2} \mid ND_2] \times P(ND_2) \\ &- \mathbb{E}[y^{0,1} - y^{2,2} \mid ID_1] \times P(ID_1) \\ &- \mathbb{E}[y^{2,1} - y^{0,0} \mid ID_2] \times P(ID_2)\end{aligned}$$

At the field level, we assume exclusion, hence:

$$\begin{aligned}\theta_1 &= \mathbb{E}[y^1 - y^0 \mid C_1] \times P(C_1) \\ &+ \mathbb{E}[y^1 - y^2 \mid C_2] \times P(C_2) \\ &+ \mathbb{E}[y^1 - y^2 \mid ND_1] \times P(ND_1) \\ &+ \mathbb{E}[y^1 - y^0 \mid ND_2] \times P(ND_2) \\ &+ \mathbb{E}[y^2 - y^0 \mid ID_1] \times P(ID_1) \\ &- \mathbb{E}[y^2 - y^0 \mid ID_2] \times P(ID_2)\end{aligned}$$

We divide by the first stage and rearrange. This gives us:

$$\begin{aligned}\tilde{\beta}_1^{\text{IV}} = &\frac{P(C_1)}{\pi_{1,0}} \underbrace{\mathbb{E}[y^1 - y^0 \mid C_1]}_{\text{A}} + \frac{P(C_2)}{\pi_{1,0}} \underbrace{\mathbb{E}[y^1 - y^2 \mid C_2]}_{\text{A}} \\ &+ \frac{P(ND_1)}{\pi_{1,0}} \underbrace{\mathbb{E}[y^1 - y^2 \mid ND_1]}_{\text{A}} + \frac{P(ND_2)}{\pi_{1,0}} \underbrace{\mathbb{E}[y^1 - y^0 \mid ND_2]}_{\text{A}} \\ &+ \frac{P(ID_1)}{\pi_{1,0}} \underbrace{\mathbb{E}[y^2 - y^0 \mid ID_1]}_{\text{B}} - \frac{P(ID_2)}{\pi_{1,0}} \underbrace{\mathbb{E}[y^2 - y^0 \mid ID_2]}_{\text{B}}\end{aligned}$$

where

$$\pi_{1,0} = P(C_1 \cup C_2 \cup ND_1 \cup ND_2)$$

This can be rewritten as:

$$\begin{aligned}\tilde{\beta}_{1,0}^{IV} = &\frac{P(C_1 \cup ND_2)}{\pi_{1,0}} \mathbb{E}[y^1 - y^0 \mid C_1 \cup ND_2] + \frac{P(C_2 \cup ND_1)}{\pi_{1,0}} \mathbb{E}[y^1 - y^2 \mid C_2 \cup ND_1] \\ &+ \frac{P(ID_1)}{\pi_{1,0}} \mathbb{E}[y^2 - y^0 \mid ID_1] - \frac{P(ID_2)}{\pi_{1,0}} \mathbb{E}[y^2 - y^0 \mid ID_2]\end{aligned}$$

C.3 Treatment Clustering

We have $S_1 = \{1,2\}$, $S_0 = \{0\}$ and seek to find an expression of the first stage, reduced form and IV estimand. We use the taxonomy in Table A3 to denote complier and defier groups.

Table A3. Taxonomy of response groups under treatment clustering.

Type	Cluster Level		Field Level			Group	
	$\tilde{d}^0$	$\tilde{d}^1$	d^0	d^1	d^2	Field	Cluster
Compliers	0	1	0		2	$\overline{C}$	C_1
	0	1	0	1			C_2
	0	1	0		1		ID_1
	0	1	0	2			ID_2
Always Takers	1	1	1		2	$\overline{AT}$	ND_1
	1	1	2	1			ND_2

Note: The table shows potential treatments for field and cluster instruments for groups impacted by the cluster instrument under treatment clustering. At the field level, d^0 indicates which treatment is taken given $Z = 0$, d^1 indicates which treatment is taken when $Z = 1$ and d^2 indicates which treatment is taken given $Z = 2$. The notation is equivalent at the cluster level. Relative to the clustered instrument, $\overline{C}$ are compliers and $\overline{AT}$ are always takers. Relative to the field instrument, C are compliers, ID are irrelevance defiers and ND are next-best defiers.

C.3.1 First Stage Applying the taxonomy to the expectation in equation (33), under field level monotonicity we get:

$$\mathbb{E}[\tilde{d}^1 - \tilde{d}^0] = P[\tilde{d}^1 - \tilde{d}^0 = 1] - P[\tilde{d}^1 - \tilde{d}^0 = -1] = P(\overline{C})$$

From equation (34) we hence have by the rank condition

$$\pi_{1,0} = P(\overline{C})$$

C.3.2 Reduced Form We use Table A3 to decompose the expectations in equation (35). Under independence and field level monotonicity, we get:

$$\begin{aligned}\theta_1 \;=\; & \mathbb{E}[\tilde{y}^{1,1} \mid \overline{C}] \times P(\overline{C}) \\ & + \mathbb{E}[\tilde{y}^{1,1} \mid \overline{AT}] \times P(\overline{AT}) \\ & - \mathbb{E}[\tilde{y}^{0,0} \mid \overline{C}] \times P(\overline{C}) \\ & - \mathbb{E}[\tilde{y}^{1,0} \mid \overline{AT}] \times P(\overline{AT})\end{aligned}$$

This rearranges to:

$$\begin{aligned}\theta_1 \;=\; & \mathbb{E}[\tilde{y}^{1,1} - \tilde{y}^{0,0} \mid \overline{C}] \times P(\overline{C}) \\ & - \mathbb{E}[\tilde{y}^{0,1} - \tilde{y}^{0,0} \mid \overline{AT}] \times P(\overline{AT})\end{aligned}$$

Using Table A2 to turn cluster level groups into field level groups, further using that groups are disjoint, and changing outcome indices to reflect instruments relevant to the

group in question, we get:

$$\begin{aligned}\theta_1 &= \mathbb{E}[y^{2,2} - y^{0,0} \mid C_1] \times P(C_1) \\ &+ \mathbb{E}[y^{1,1} - y^{0,0} \mid C_2] \times P(C_2) \\ &+ \mathbb{E}[y^{1,2} - y^{0,0} \mid ID_1] \times P(ID_1) \\ &+ \mathbb{E}[y^{2,1} - y^{0,0} \mid ID_2] \times P(ID_2) \\ &- \mathbb{E}[y^{2,2} - y^{1,0} \mid ND_1] \times P(ND_1) \\ &- \mathbb{E}[y^{1,1} - y^{2,0} \mid ND_2] \times P(ND_2)\end{aligned}$$

At the field level, we assume exclusion, hence:

$$\begin{aligned}\theta_1 &= \mathbb{E}[y^2 - y^0 \mid C_1] \times P(C_1) \\ &+ \mathbb{E}[y^1 - y^0 \mid C_2] \times P(C_2) \\ &+ \mathbb{E}[y^1 - y^0 \mid ID_1] \times P(ID_1) \\ &+ \mathbb{E}[y^2 - y^0 \mid ID_2] \times P(ID_2) \\ &- \mathbb{E}[y^2 - y^1 \mid ND_1] \times P(ND_1) \\ &- \mathbb{E}[y^1 - y^2 \mid ND_2] \times P(ND_2)\end{aligned}$$

We divide by the first stage and rearrange. This gives us:

$$\begin{aligned}\tilde{\beta}_1^{\text{IV}} &= \frac{P(C_1)}{\pi_{1,0}}\ \mathbb{E}[y^2 - y^0 \mid C_1]\ +\ \frac{P(C_2)}{\pi_{1,0}}\ \mathbb{E}[y^1 - y^0 \mid C_2] \\ &+ \frac{P(ID_1)}{\pi_{1,0}}\ \mathbb{E}[y^1 - y^0 \mid ID_1]\ +\ \frac{P(ID_2)}{\pi_{1,0}}\ \mathbb{E}[y^2 - y^0 \mid ID_2] \\ &+ \frac{P(ND_1)}{\pi_{1,0}}\ \mathbb{E}[y^1 - y^2 \mid ND_1]\ -\ \frac{P(ND_2)}{\pi_{1,0}}\ \mathbb{E}[y^1 - y^2 \mid ND_2]\end{aligned}$$

where

$$\pi_{1,0} = P(C_1 \cup C_2 \cup ID_1 \cup ID_2)$$

This may be rewritten to

$$\begin{aligned}\tilde{\beta}_{1,0}^{IV} &= \frac{P(C_1 \cup ID_1)}{\pi_{1,0}}\mathbb{E}[y^1 - y^0 \mid C_1 \cup ID_1] + \frac{P(C_2 \cup ID_1)}{\pi_{1,0}}\mathbb{E}[y^2 - y^0 \mid C_2 \cup ID_1] \\ &+ \frac{P(ND_1)}{\pi_{1,0}}\ \mathbb{E}[y^1 - y^2 \mid ND_1] - \frac{P(ND_2)}{\pi_{1,0}}\ \mathbb{E}[y^1 - y^2 \mid ND_2]\end{aligned}$$

D Examining violations of next-best and irrelevance and payoffs

Table A4. Joint test of irrelevance and next-best by completed field, Norway

	on-diagonal (α^j_{jl})			off-diagonal ($\alpha^k_{jl}, j \notin \{k,l\}$)		
Completed field	first stages	F-statistic	p-value	first stages	F-statistic	p-value
Science	7	28.2	<0.01	49	4.7	<0.01
Business	7	42.7	<0.01	49	7.6	<0.01
Social Science	7	26.2	<0.01	49	8.3	<0.01
Teaching	7	124.0	<0.01	49	5.2	<0.01
Humanities	7	20.6	<0.01	49	5.9	<0.01
Other Health	7	420.8	<0.01	49	4.7	<0.01
Technology	7	50.6	<0.01	49	4.3	<0.01
Law	7	94.9	<0.01	49	4.9	<0.01
Medicine	8	93.6	<0.01	56	5.7	<0.01
All	*64*	*100.65*	*<0.01*	*448*	*10.57*	*<0.01*

Table A5. Off-diagonal first stages by completed field, sign and significance, Norway

	off-diagonal, # firsts stages that are					
	>0			<0		
Completed field	All	Sign.	Multi sign.	All	Sign.	Multi sign.
Science	18	2		31	17	6
Business	5			44	21	8
Social Science	4	1		45	24	12
Teaching	37	13		12	6	5
Humanities	12	3	1	37	15	8
Other Health	30	9		19	9	6
Technology	11	1		38	14	5
Law	11			38	18	4
Medicine	18	1		38	20	7
All	*146*	*30*	*1*	*302*	*144*	*61*

Note: Significant are first stages with $p < .05$, multi test-significant are first stages with $p < .05/512$ (i.e. significant with a Bonferroni adjustment for multiple testing).

Table A6. Joint test of irrelevance and next-best by completed field, Denmark

	on-diagonal (α_{jl}^{j})			off-diagonal ($\alpha_{jl}^{k}, j \notin \{k,l\}$)		
Completed field	first stages	F-statistic	p-value	first stages	F-statistic	p-value
Science	7	95.3	<0.01	49	2.1	<0.01
Business	7	2.8	<0.01	49	1.9	<0.01
Social Science	7	111.5	<0.01	49	3.9	<0.01
Teaching	7	12.7	<0.01	49	2.6	<0.01
Humanities	7	33.7	<0.01	49	1.8	<0.01
Other Health	7	61.5	<0.01	49	2.8	<0.01
Technology	7	38.2	<0.01	49	2.2	<0.01
Law	7	63.3	<0.01	49	1.4	0.04
Medicine	8	98.8	<0.01	56	1.6	<0.01
All	*64*	*57.9*	*<0.01*	*448*	*3.3*	*<0.01*

Table A7. Off-diagonal first stages by completed field, sign and significance, Denmark

	off-diagonal, # firsts stages that are					
	>0			<0		
Completed field	All	Sign.	Multi sign.	All	Sign.	Multi sign.
Science	26	7		23	8	1
Business	23	5	1	26	6	1
Social Science	29	8	2	20	10	5
Teaching	30	4		19	4	1
Humanities	32	8		17	2	
Other Health	30	7	3	19	2	
Technology	26	2		23	7	
Law	22	2		27	6	
Medicine	24	2		32	13	1
Sum	*242*	*45*	*6*	*206*	*58*	*9*

Note: Significant are first stages with $p < .05$, multi test-significant are first stages with $p < .05/512$ (i.e. significant with Bonferroni adjustment for multiple testing).

E The Danish institutional setting and implications for empirical specification

Danish admission institutions For programs with restricted admission, student places are allocated through two quotas.

The majority of places are allocated through Quota 1 based on applicants' GPA from high school, although some programs have additional specific requirements, e.g., a high-level maths course from high school. In our data period grades were awarded on a 10-point scale with integer values between 0 and 13 (omitting values 1, 2, 4 and 12). The high school GPA is based on grades in all the subjects on the student's study program, and it is recorded to 1 decimal place. All applicants with a GPA strictly above the Quota 1 threshold level are admitted provided they also meet any specific entry requirements. Because of the coarse GPA measure, there will often not be sufficient places for all applicants with a GPA exactly equal to the threshold, and in this case the oldest are typically admitted first. In our sample, about one third of the applicants with a GPA exactly equal to the threshold are not admitted to their preferred program. It is important to note that the minimum GPAs needed to be admitted (the GPA thresholds) are published after the student places are allocated and that the variation in thresholds over time is considerable, and thus applicants cannot predict the exact thresholds.

Most programs also have a standby (waiting) list and the GPA threshold for the standby list is typically a little lower than the Quota 1 threshold. On the application form, applicants can choose to apply for the standby list as well. If some of the applicants admitted under Quota 1 drop out before the course starts or in the very early days of the course, then their places are offered to applicants on the standby list. In any event, applicants admitted to the standby list are guaranteed a study place the following year. Applicants who are admitted to a standby list are not considered for any of the lower-ranked programs on their application.

Some of the available student places are reserved for admission via Quota 2 where applicants are assessed based on other criteria besides their GPA from upper secondary school, e.g. specific admission tests, admission interviews, or vocational qualifications. The institutions have considerable discretion in deciding these criteria and the relative size of Quotas 1 and 2 for their programs. Quota 2 applicants are automatically considered for Quota 1, so if they meet the Quota 1 GPA criterion (and any additional specific criteria), they are admitted via Quota 1.

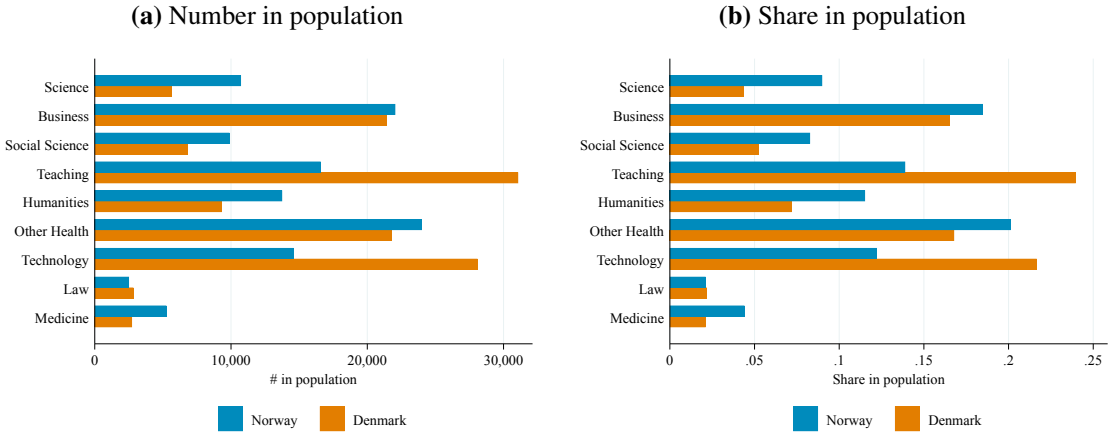
We do not observe whether students apply for Quota 2 or standby, but for those who are offered a program, we know whether admission was via Quota 1, standby or Quota 2.

Specification of instrument The coarse application score variable and the standby list option are handled by using the following instrument:

$$z = [GPA > Q1] + [Std \leq GPA \leq Q1] \times \text{Offer}$$

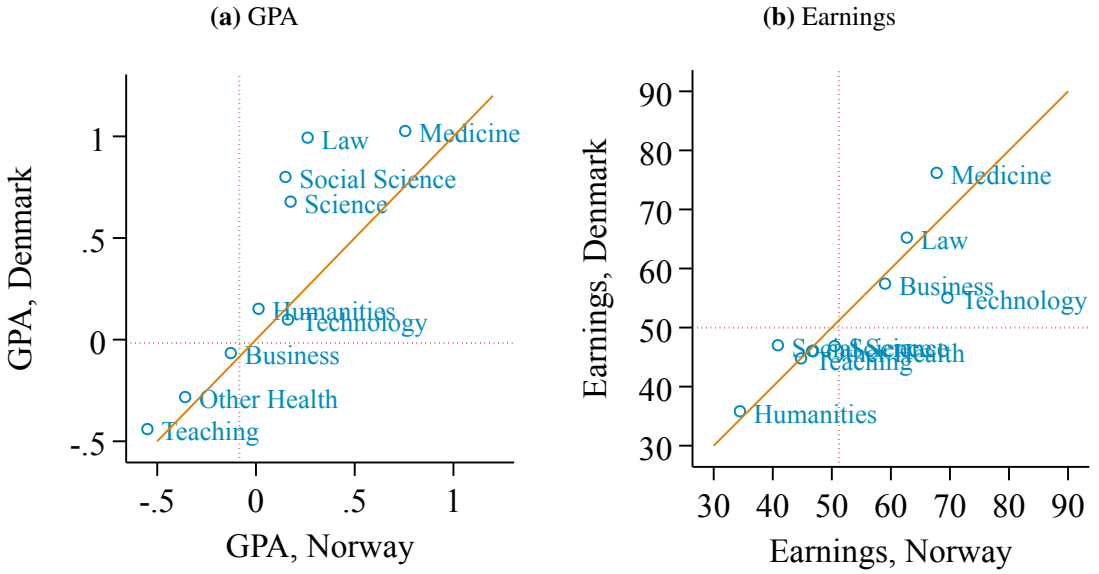
where $Q1$ and Stb are the Quota 1 and standby admission thresholds of the preferred field of the applicant, respectively. The first term on the right-hand side captures the effect of having an application score strictly above the Quota 1 threshold to the preferred field k. When $GPA = Q1$ or $GPA = Std$ the offer is a random draw (conditional on age which is controlled for in the analysis, and for the standby threshold also conditional on having applied for standby), and $[GPA \in \{Q1, Std\}] \times \text{Offer}$ is therefore a valid instrument. When $[Std < GPA < Q1]$ those who applied to the standby list will receive an offer.

F Extra results



Note: Figures show the distribution of completed fields in the population for those born 1979-1983 (Norway) or 1975-81 (Denmark) that have completed any higher education. Completed field is measured at age 28.

Figure A1. Distribution of completed field in the population



Note: Figure shows population-weighted average GPA of applicants and population earnings, based on the populations in Figure A1. GPA is demeaned within country, but not otherwise standardized. Earnings are measured at age 28.

Figure A2. GPA and earnings in Norway and Denmark, average by country and field

Table A8. Payoff estimates Norway ($1,000), 8 years after application

	Next-best field							
	Science	Business	Social Science	Teaching	Humanities	Other Health	Technology	Law
Completed field								
Science		-22.9	27.0	4.8	8.4	4.0	-17.3	145.0
		(17.9)	(25.8)	(20.1)	(15.6)	(15.5)	(16.3)	(167.1)
Business	70.8		59.3	26.3	41.6	26.9	-0.0	24.0
	(10.5)		(9.5)	(6.6)	(7.9)	(9.0)	(6.1)	(32.5)
Social Science	63.9	-33.6		5.0	13.4	-11.1	-45.4	-48.5
	(18.6)	(10.3)		(10.5)	(6.9)	(12.4)	(23.3)	(53.4)
Teaching	47.3	-21.3	32.3		19.3	0.4	-29.6	7.8
	(9.7)	(5.5)	(6.9)		(4.4)	(4.8)	(6.7)	(46.6)
Humanities	8.2	-36.2	24.7	-5.1		-22.5	-37.7	-87.7
	(15.2)	(8.5)	(9.9)	(7.9)		(11.3)	(7.8)	(87.6)
Other Health	47.5	-17.1	32.2	6.3	17.0		-26.0	-28.2
	(9.5)	(3.8)	(6.8)	(2.3)	(4.8)		(4.7)	(39.9)
Technology	87.4	-5.6	66.1	35.8	57.9	27.8		-20.5
	(11.9)	(7.1)	(9.5)	(8.0)	(9.0)	(8.3)		(57.0)
Law	52.2	-11.9	53.7	28.8	40.3	22.3	-18.6	
	(11.0)	(7.9)	(7.5)	(10.6)	(5.9)	(10.7)	(9.6)	
Medicine	97.9	20.4	76.6	55.9	74.3	42.1	20.5	42.1
	(11.3)	(9.1)	(9.4)	(8.5)	(8.3)	(6.6)	(5.8)	(28.8)
N	5,320	4,477	11,250	11,254	8,539	3,371	3,612	1,403

Note: 2SLS estimation of equations (11)–12 results in a matrix of payoffs to field j as compared to k for those who prefer j and have k as next-best field. Each cell is a 2SLS estimate (with standard errors in parenthesis) of the payoff to a given pair of preferred field and next-best field. The rows represent completed fields and the columns represent next-best fields.

Table A9. Payoff estimates Denmark ($1,000), 8 years after application

	Next-best field							
	Science	Business	Social Science	Teaching	Humanities	Other Health	Technology	Law
Completed field								
Science		-10.8	-10.4	-4.5	0.5	5.6	-23.5	-5.8
		(9.6)	(10.0)	(16.8)	(10.0)	(7.9)	(9.4)	(6.7)
Business	61.7		18.3	31.7	54.7	25.5	33.6	16.6
	(22.9)		(17.4)	(26.2)	(20.5)	(29.6)	(23.3)	(18.6)
Social Science	-15.2	-23.3		-4.3	15.5	-4.4	-8.3	-8.0
	(10.2)	(8.1)		(4.3)	(5.0)	(4.4)	(8.6)	(4.0)
Teaching	-9.0	-35.5	-16.9		-0.5	-21.2	-23.0	-16.9
	(10.3)	(11.1)	(8.1)		(8.8)	(8.3)	(8.4)	(9.5)
Humanities	-9.6	-32.2	-31.3	-9.5		-24.1	-28.9	-31.4
	(8.3)	(7.7)	(10.0)	(5.8)		(10.7)	(10.9)	(9.8)
Other Health	15.0	-24.7	-3.7	-3.6	16.4		-18.6	-20.2
	(10.6)	(9.3)	(7.2)	(2.9)	(5.3)		(8.9)	(6.8)
Technology	7.1	-17.4	-13.7	-17.2	9.9	-7.8		-4.9
	(7.9)	(9.5)	(9.6)	(19.1)	(7.9)	(8.8)		(16.6)
Law	22.4	-4.1	12.4	1.5	31.9	16.6	9.3	
	(6.2)	(5.2)	(6.3)	(8.8)	(5.6)	(4.7)	(8.6)	
Medicine	21.7	5.4	7.1	29.4	28.9	11.9	13.7	14.1
	(5.1)	(8.2)	(6.9)	(8.0)	(6.1)	(4.3)	(4.5)	(5.9)
Total	1,254	1,076	1,427	1,359	2,761	1,521	738	432

Note: 2SLS estimation of equations (11)–12 results in a matrix of payoffs to field j as compared to k for those who prefer j and have k as next-best field. Each cell is a 2SLS estimate (with standard errors in parenthesis) of the payoff to a given pair of preferred field and next-best field. The rows represent completed fields and the columns represent next-best fields.

Table A10. Payoff estimates Norway ($1,000), 13 years after application

	Next-best field							
	Science	Business	Social Science	Teaching	Humanities	Other Health	Technology	Law
Completed field								
Science		-15.1	-3.2	23.1	-3.1	2.1	-19.5	-10.3
		(22.0)	(29.6)	(27.9)	(20.6)	(22.8)	(19.7)	(185.7)
Business	52.8		36.9	36.6	32.8	30.5	1.6	-0.8
	(18.5)		(16.6)	(12.0)	(14.8)	(13.2)	(9.9)	(63.1)
Social Science	55.5	-46.2		17.6	12.5	-8.6	-70.2	-18.7
	(26.8)	(15.9)		(15.6)	(10.2)	(19.5)	(29.2)	(71.3)
Teaching	19.5	-35.8	5.1		4.4	-1.3	-46.0	9.8
	(16.0)	(8.9)	(10.9)		(6.4)	(7.6)	(10.4)	(189.2)
Humanities	-5.3	-36.6	12.8	12.4		-3.7	-36.6	-182.4
	(21.8)	(12.9)	(15.6)	(10.9)		(12.5)	(10.8)	(495.3)
Other Health	9.9	-36.0	1.3	3.3	-2.2		-45.7	-55.2
	(15.8)	(4.9)	(11.3)	(3.3)	(7.9)		(6.3)	(56.7)
Technology	62.0	-6.2	57.8	49.5	49.3	56.3		-20.9
	(19.1)	(9.2)	(16.0)	(11.4)	(12.1)	(19.2)		(49.2)
Law	33.4	-19.0	37.1	37.3	38.6	37.7	-11.7	
	(16.4)	(10.6)	(12.8)	(9.5)	(9.4)	(15.4)	(11.9)	
Medicine	67.0	22.2	54.2	76.4	64.0	57.5	21.1	47.4
	(20.1)	(20.6)	(17.3)	(19.0)	(12.9)	(12.3)	(8.7)	(67.8)
N	5,005	4,402	11,010	11,120	8,293	3,284	3,480	1,362

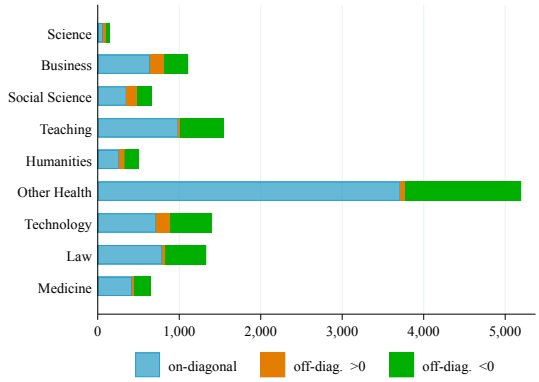
Note: 2SLS estimation of equations (11)–12 results in a matrix of payoffs to field j as compared to k for those who prefer j and have k as next-best field. Each cell is a 2SLS estimate (with standard errors in parenthesis) of the payoff to a given pair of preferred field and next-best field. The rows represent completed fields and the columns represent next-best fields.

Table A11. Payoff estimates Denmark ($1,000), 13 years after application

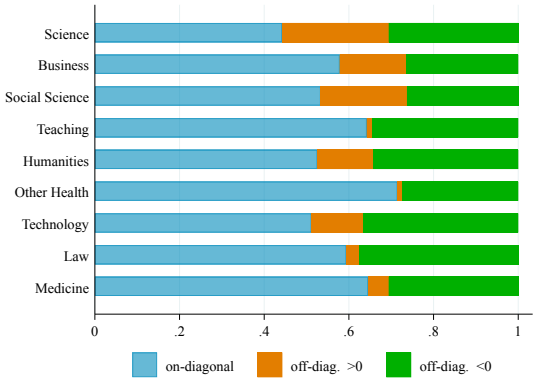
	Next-best field							
	Science	Business	Social Science	Teaching	Humanities	Other Health	Technology	Law
Completed field								
Science		-17.6	19.8	33.7	0.7	-2.7	-24.3	-24.3
		(13.5)	(23.4)	(15.9)	(10.1)	(9.2)	(12.1)	(11.9)
Business	75.0		42.7	76.9	81.4	54.7	74.1	47.0
	(27.4)		(24.1)	(36.6)	(29.9)	(32.7)	(28.6)	(21.9)
Social Science	-15.2	-22.4		0.6	12.8	4.7	-15.8	-16.5
	(11.6)	(12.8)		(5.0)	(5.6)	(5.9)	(9.8)	(5.4)
Teaching	-31.5	-43.1	-35.0		-8.3	-23.1	-34.8	-40.6
	(12.4)	(13.3)	(10.5)		(10.9)	(9.8)	(11.9)	(11.4)
Humanities	-29.1	-21.0	-20.6	-0.1		-0.4	-22.4	-31.2
	(9.5)	(10.1)	(12.5)	(5.9)		(11.3)	(12.0)	(10.3)
Other Health	-13.7	-35.9	-12.4	-5.9	10.0		-13.3	-49.4
	(11.9)	(14.8)	(10.9)	(3.8)	(6.7)		(10.7)	(14.1)
Technology	-4.5	-45.5	-39.3	-19.8	3.5	-12.0		-25.5
	(10.6)	(13.0)	(12.7)	(19.8)	(9.2)	(15.8)		(17.6)
Law	22.7	3.4	12.5	21.8	30.1	16.3	25.1	
	(8.1)	(7.6)	(8.9)	(8.7)	(7.2)	(8.7)	(12.1)	
Medicine	14.6	7.8	14.5	46.1	27.1	30.5	13.7	2.4
	(6.5)	(11.2)	(9.0)	(15.6)	(6.7)	(5.1)	(5.5)	(7.2)
N	1,477	1,197	1,701	1,493	3,324	1,667	824	520

Note: 2SLS estimation of equations (11)–12 results in a matrix of payoffs to field j as compared to k for those who prefer j and have k as next-best field. Each cell is a 2SLS estimate (with standard errors in parenthesis) of the payoff to a given pair of preferred field and next-best field. The rows represent completed fields and the columns represent next-best fields.

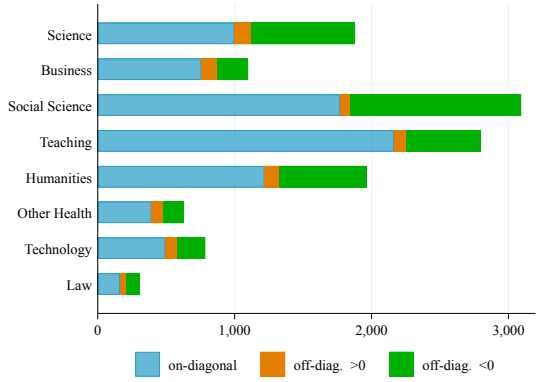
(a) Numbers of applicants shifted by preferred field



(b) Share of applicants shifted by preferred field



(c) Numbers of applicants shifted by next-best field



(d) Share of applicants shifted by next-best field

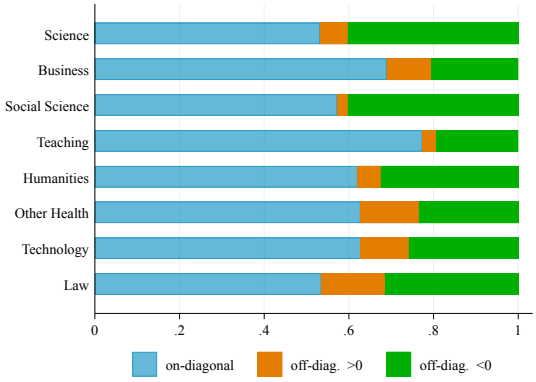
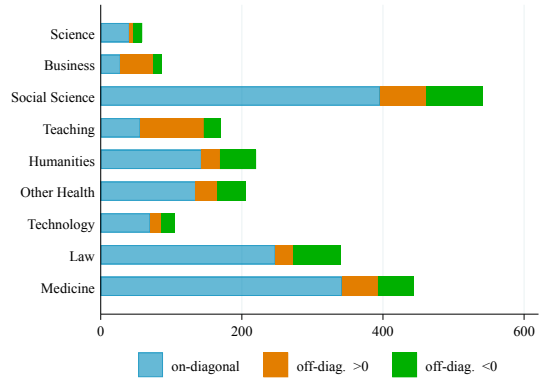
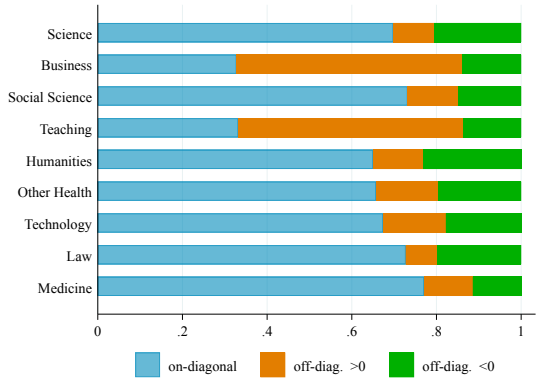


Figure A3. Numbers and shares of applicants shifted by the instrument by violating irrelevance or not and by preferred/stated next-best field, Norway

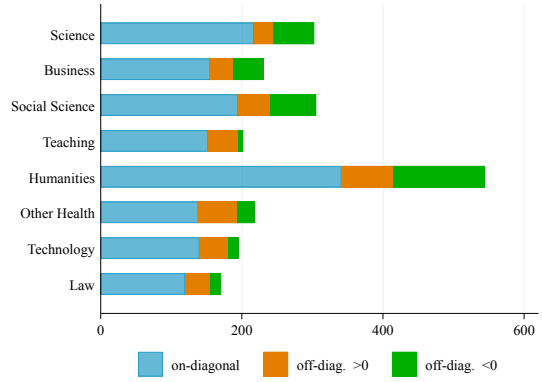
(a) Numbers of applicants shifted by preferred field



(b) Share of applicants shifted by preferred field



(c) Numbers of applicants shifted by next-best field



(d) Share of applicants shifted by next-best field

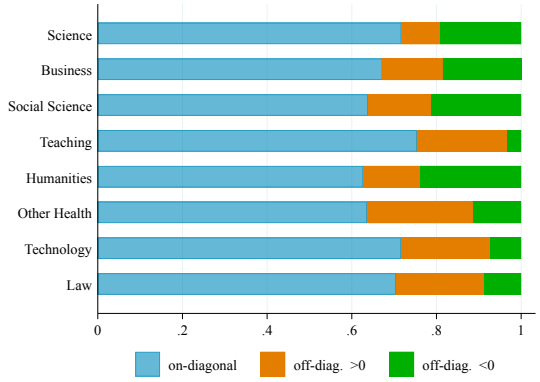


Figure A4. Numbers and shares of applicants shifted by the instrument by violating irrelevance or not and by preferred/stated next-best field, Denmark

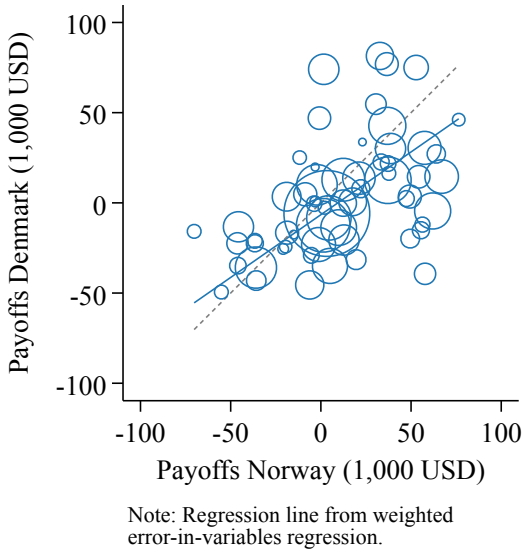


Figure A5. Payoffs in Norway and Denmark 13 years after applying, all completed and next-best fields