# The causal effect of sponsoring foreign workers on UK firms' productivity

**Dimitris Christelis**
University of Glasgow

**Ali Gokhan Karabilgin**
University of Glasgow

# Motivation

- Slowdown in productivity growth in the UK
- Big debate about the effects of immigration
- Do immigrant workers improve the productivity of UK firms?

# The link between foreign workers and productivity

- Labour productivity (LP):
  - fills hard-to-fill specialist roles and reduces vacancy or project delays
  - improves worker-job matching, task allocation, and team complementarities
  - may support scale expansion and capital deepening

- Total factor productivity (TFP):
  - transfers technical, managerial, and foreign-market knowledge
  - facilitates intra-company mobility and organisational upgrading

# The link between foreign workers and productivity

- Potential gains:
  - scarce skills, better matches, knowledge transfer, lower vacancy costs
  - stronger complementarities and retention of firm-specific human capital

- Potential costs and frictions:
  - licence, CoS, Immigration Skills Charge, HR and legal costs, hiring delays and reduced flexibility
  - tied-worker mobility frictions may generate rents rather than technical efficiency

# Causality

**Causal statements are difficult in this context:**

- sponsor licence holding is a firm choice
- firms apply when they expect access to regulated migrant labour to be valuable
- sponsors may already be better managed, larger, more skill-intensive, and more internationally oriented
- anticipated demand or productivity gains may affect current sponsorship
- Foreign workers may self-select into more productive firms

# Causality

- Panel methods: remove time-invariant heterogeneity, but not time-varying unobservables
- Local migration / shift-share IVs: operate at the local-labour-market level and require difficult exclusion restrictions for a firm-level sponsorship decision. Moreover, only LATE estimated
- Diff-in-diff / event studies: require parallel trends, clean adoption timing, limited anticipation, and strict exogeneity
- Measurement: realised firm-level sponsorship is unobserved; licence status measures access rather than use

# Literature

- Regional and sectoral studies often report positive effects or associations using variation in migrant shares across states, regions, sectors, or local labour markets (Peri, 2012; Campo et al., 2018; Mitaritonna et al., 2017; Nam and Portes, 2023)

- Firm-level studies report gains in UK service firms, Danish firms hiring foreign experts, and UK firms hiring foreign managers (Ottaviano et al., 2018; Malchow-Møller et al., 2019; Exadaktylos et al., 2024)

- Other evidence finds weak or heterogeneous effects and emphasises migrant sorting across firms (Paserman, 2013; Trax et al., 2015; Hall and Manning, 2024)

- Existing evidence suggests potential productivity gains, but more robust evidence needed at the firm level, and for population-level results

# Literature – employer-driven migration

- H-1B lottery evidence finds modest innovation effects and substantial crowd-out of other workers at the margin (Doran et al., 2022)
- Restrictions on high-skilled visas can induce multinational firms to shift activity abroad rather than replace foreign workers one-for-one with domestic workers (Glennon, 2024)
- Foreign experts may raise productivity and exports through training, knowledge transfer, and complementarities (Markusen and Trofimenko, 2009; Malchow-Møller et al., 2019)
- Qualitative UK evidence suggests that Tier 2 recruitment can support workforce flexibility and project delivery while generating compliance and adjustment costs (Rolfe et al., 2013)

# Contribution

- We link Home Office sponsor registers to the FAME firm database and identify firm-year participation in the UK employer-sponsored migration system

- To address the issue of causality we use partial identification methods that bound the effect of foreign worker sponsorship on productivity using mild, and thus credible, assumptions that do not rely on any exogenous variation

# Data

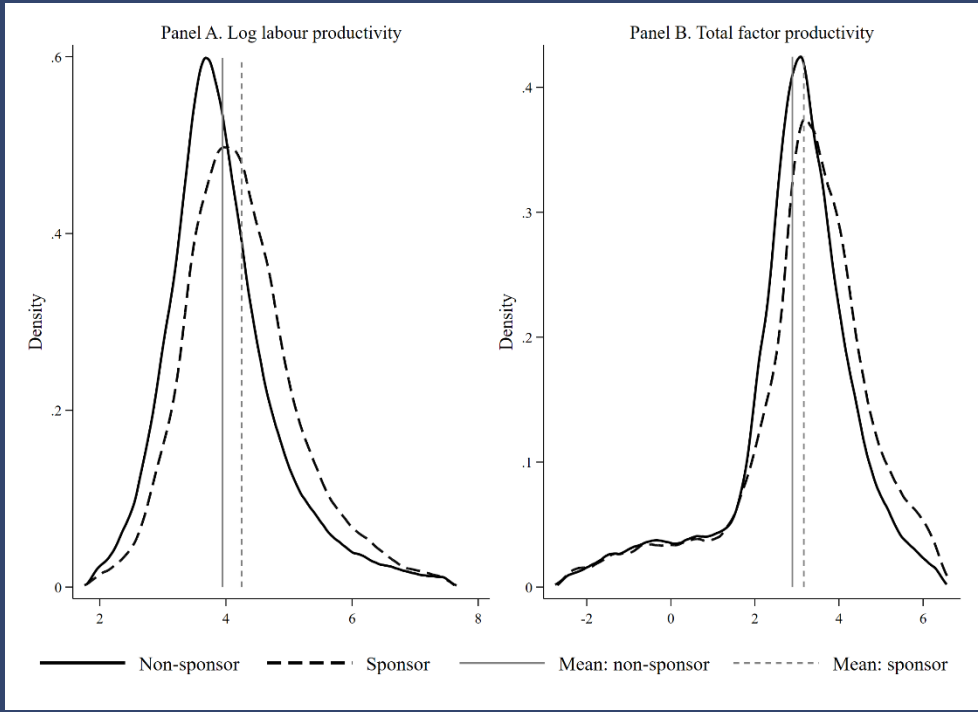
- FAME database – standard source of firm data for the UK
- Balance sheet and profit loss statements available
- Output: gross value added
- Labor input: number of employees (not hours)
- We exclude agriculture, finance, public sector firms, and the household sector
- TFP is calculated using the methodology in Ackeberg et al (2015, a modification of Olley and Pakes, 1996)

# Data (cont.)

- Archived UK Border Agency and Home Office sponsor registers
- Sponsor organisations linked to FAME through Companies House registered numbers
- Main estimation window: 2017–2019:
  after the Immigration Skills Charge; before the end-of-free-movement reforms – EU and long-term non-EU UK residents do not need sponsorship
- LP sample:
  177,161 firm-years; 27,114 sponsor observations; 11,191 sponsor firms
- TFP sample:
  105,397 firm-years; 18,324 sponsor observations; 7,986 sponsor firms

# Data (cont.)

## Distribution of firm productivity by sponsor status




Mean log LP:

- 4.25 sponsors
- 3.95 non-sponsors

Mean TFP:

- 3.18 sponsors
- 2.91 non-sponsors

# Data (cont.)

**Sponsor vs non-sponsor means:**

- Mean no of mployees: 933 vs 377
- Exporter: 52.9% vs 37.2%
- Real remuneration per employee: £72.3k vs £45.7k
- Lagged tangible capital per employee: £129k vs £134k
- Return on tangible assets: not much difference
- Sponsors are larger, higher-paying, and more international

# Partial identification

- We use partial identification methods (Manski, 1990; Manski, 1997; Manski and Pepper, 2000)

- Starting premise:
  - the treatment is endogenous (any form of endogeneity can be accommodated) and there is no exogenous variation that can help point identify the ATE/ATT

- PI puts bounds around counterfactual outcomes, and thus around the ATE/ATT, using credible assumptions

# Partial identification

- Effect of interest: average treatment effect on the treated (outcome $Y$, observed treatment $w$, binary sponsorship treatment, controls $x$) :

$$E[Y(1)|x,\ w=1]-E[Y(0)\ |x,\ w=1]$$

# Partial identification (cont.)

- Observed sample analogue of $E[Y(1)|x,w=1]$
- Unobserved counterfactual $E[Y(0)\,|\,x,w=1]$
- Exogeneity implies $E[Y(0)\,|\,x,w=1]=E[Y(0)|x,w=0]$. Thus, the ATT becomes

$$E[Y(1)|x,\ w=1]-E[Y(0)\,|x,\ w=1]=$$
$$E[Y(1)|x,\ w=1]-E[Y(0)|x,\ w=0]$$

# Problem of the counterfactual

- Two standard approaches:
  - statistical models, typically regression-based
  - models modelling explicitly decision-making
- Both approaches rely on exogeneity
  - error distributions are assumed to have no dependence on the treatment, observables, or other parameters in the model
  - the same estimating equation/decision rule can be applied irrespective of actual treatment received
- Intuition: all systematic factors are incorporated in the statistical or optimization model; all that is left out of the model is unsystematic
  - Very strong assumption: thousands of firms in different sectors that almost surely differ in various important unobservables, including managerial qualities, market knowledge, supply chains, foreign connections etc.

# Partial identification

- PI postulates that one cannot included all systematic factors in the model; hence exogeneity cannot be assumed, which leads to $E[Y(0) \mid x, w=1] \neq E[Y(0) \mid x, w=0]$
- Instead of assuming $E[Y(0) \mid x, w=1] = E[Y(0) \mid x, w=0]$, put bounds around $E[Y(0) \mid x, w=1]$ using observable quantities
- Bounds on $E[Y(0) \mid x, w=1]$ imply bounds on average potential outcomes, and bounds on the differences in average potential outcomes, that is on the ATE/ATT
- PI is non-parametric, relies on few and credible assumptions, and is completely transparent about how each assumption affects results. Trade-off: credibility versus uncertainty

# Partial Identification (cont.)

- For example,

$$E[Y(1)|x, w{=}1] = 5$$
$$2 \leq E[Y(0)|x, w{=}1] \leq 4$$

Then,

$$5 - 4 = 1$$
$$\leq E[Y(1)|x, w{=}1] - E[Y(1)|x, w{=}0)] \leq$$
$$5 - 2 = 3$$

- To narrow identification regions, we need lower bounds that are as large as possible, and upper bounds that are as small as possible

# Partial Identification (cont.)

- Without assumptions, the only credible bounds on the counterfactual terms are the minimum and maximum possible values, which typically produce uninformative identification regions

$$E[Y(d)|x\,,\,w{=}1] - y_{max}$$
$$\leq E[Y(1)|x\,,\,w{=}1] - E[Y(0)|x\,,\,w{=}1] \leq$$
$$E[Y(d)|x\,,\,w{=}1] - y_{min}$$

- Using assumptions, we will replace $y_{min}$ with something larger, and $y_{max}$ with something smaller

# Assumption 1: Monotone Treatment Response (Manski, 1997)

- Foreign sponsorship does not decrease productivity, ***on average***, ***keeping everything else constant.*** For $d_2 > d_1$

$$E[Y(d_2)|w = 1] \geq E[Y(d_1)|w = 1)]$$

- Weak monotonicity, thus equality allowed - not sufficient to exclude zero from the ATT identification region

- Plausible assumption – difficult to think why widening the pool of worker talent would reduces productivity, on average.

- How it works:

$$E[Y(0)|w = 1)] \leq E[Y(1)|w = 1] \leq y_{max}$$

# Assumption 2: Monotone Treatment Selection (Manski and Pepper, 2000)

- Firms that sponsor foreign workers have weakly higher productivity than firms that do not for any common level of sponsorship status, ***on average.*** For $\forall d$

$$E[Y(d)|w = 1] \geq E[Y(d)|w = 0)]$$

- Weak monotonicity, thus equality allowed. MTS characterises the direction of treatment endogeneity

- Plausible assumption – firms that sponsor foreign workers likely to have better opportunities to use the latter's talent and more far-sighted management – on average.

- How it works:

$$E[Y(0)|w = 1)] \geq E[Y(0)|w = 0] \geq y_{min}$$

# Joint test of MTR+MTS (Manski and Pepper, 2000)

- Very important to test as many assumptions as possible for the credibility of results

- Manski and Pepper (2000) show that MTR+MTS implies that

$$E[Y(1)|w = 1] \geq E[Y(1)|w = 0] \geq E[Y(0)|w = 0)]$$

The first inequality holds due to MTS, and the second due to MTR.

- Very easy to test this condition regarding weak monotonicity of observed conditional mean outcomes, which clearly holds in our data

# Assumption 3: Monotone Instrumental Variable (Manski and Pepper, 2000)

- MIVs can be weakly monotonically associated with the outcome, and are thus much easier to find
- They do not identify the LATE but the ATE – nothing to do with compliers
- An instrument $z$ is monotone if

$$z_1 < z_2 \Rightarrow E[Y(d)|Z = z_1] \leq E[Y(d)|Z = z_2]$$

- Assumption is about potential outcome $E[Y(d)]$, hence it is untestable. It also reflects a simple correlation, not a causal relationship

# Monotone instrumental variable (MIV)

- MIV condition allows a directed search in the sample space, through conditioning on the MIV, for larger lower bounds and smaller upper bounds

- Our MIVs: K/L and EBITDA/K
  - Assumption: given sponsorship status, the higher K/L is, the weakly higher current LP is. Similarly for EBITDA/K and TFP

- Weak assumption, denotes only a weakly positive association, on average, between K/L and EBITDA/K with productivity

# Partial Identification Results

| | Labour Productivity | | | Total Factor Productivity | |
|---|---|---|---|---|---|
| | Lower Bound | Upper Bound | | Lower Bound | Upper Bound |
| Exogenous Treatment Selection | 0.361 | | | 0.272 | |
| No assumptions | -3.293 | 2.420 | | -3.307 | 5.771 |
| MTR | 0.000 | 2.420 | | 0.000 | 5.771 |
| MTS | -3.293 | 0.361 | | -3.307 | 0.272 |
| MTR–MTS | 0.000 | 0.361 | | 0.000 | 0.272 |
| MIV | -3.293 | 2.420 | | -3.307 | 5.771 |
| MTR–MIV | 0.032 | 2.420 | | 0.012 | 5.771 |
| MTS-MIV | -3.293 | 0.201 | | -3.307 | 0.145 |
| MTR-MTS-MIV | 0.032 | 0.201 | | 0.012 | 0.145 |
| | | | | | |
| | | | | | |

# Conclusions

- We find a positive and economically substantial causal effect of sponsoring firm workers on both firm LP and firm TFP using micro data from FAME
- Results are obtained via PI methods using weak, and thus credible, assumptions
- Observational data (even cross-sectional ones) can be used to make causal statements, without the use of exogenous variation
- Future work: quantile treatment effects, expanding the sample to post-Brexit years

# THANK YOU!

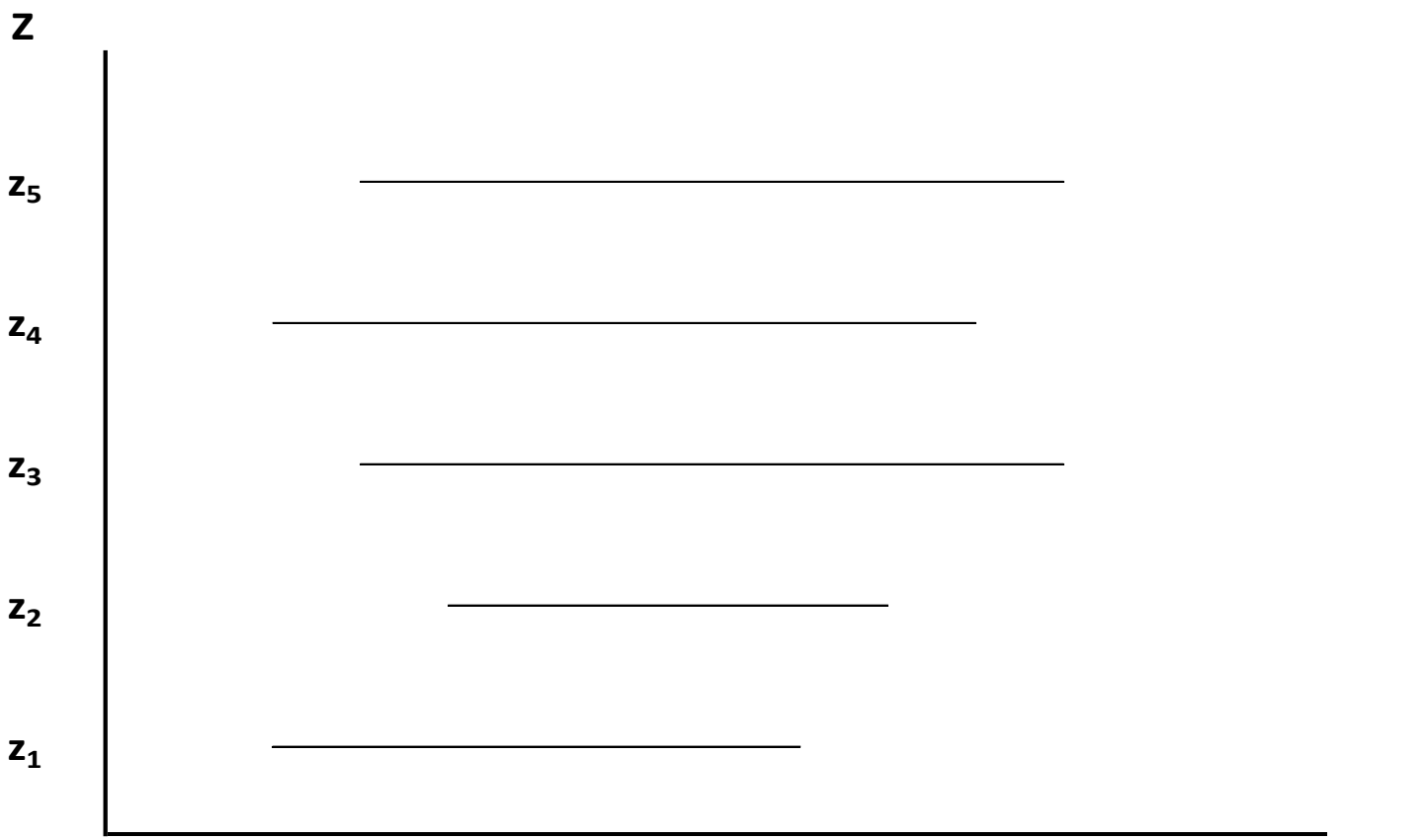
# How do MIVs work? - XIVs

- Let us assume for the moment that the instrument is exogenous (XIV)

$$E[Y(d)|Z = z] = E[Y(d)], \forall\, z$$

- Need to narrow the identification region
  - LB example: $LB(d) = \sum_z LB(d|Z = z)P(Z = z)$
  - need to increase $LB(d|Z = z)$, as $P(Z = z)$ is given by the data
  - Correspondingly, need to decrease $UB(d|Z = z)$

# How do MIVs work? - XIVs (cont.)

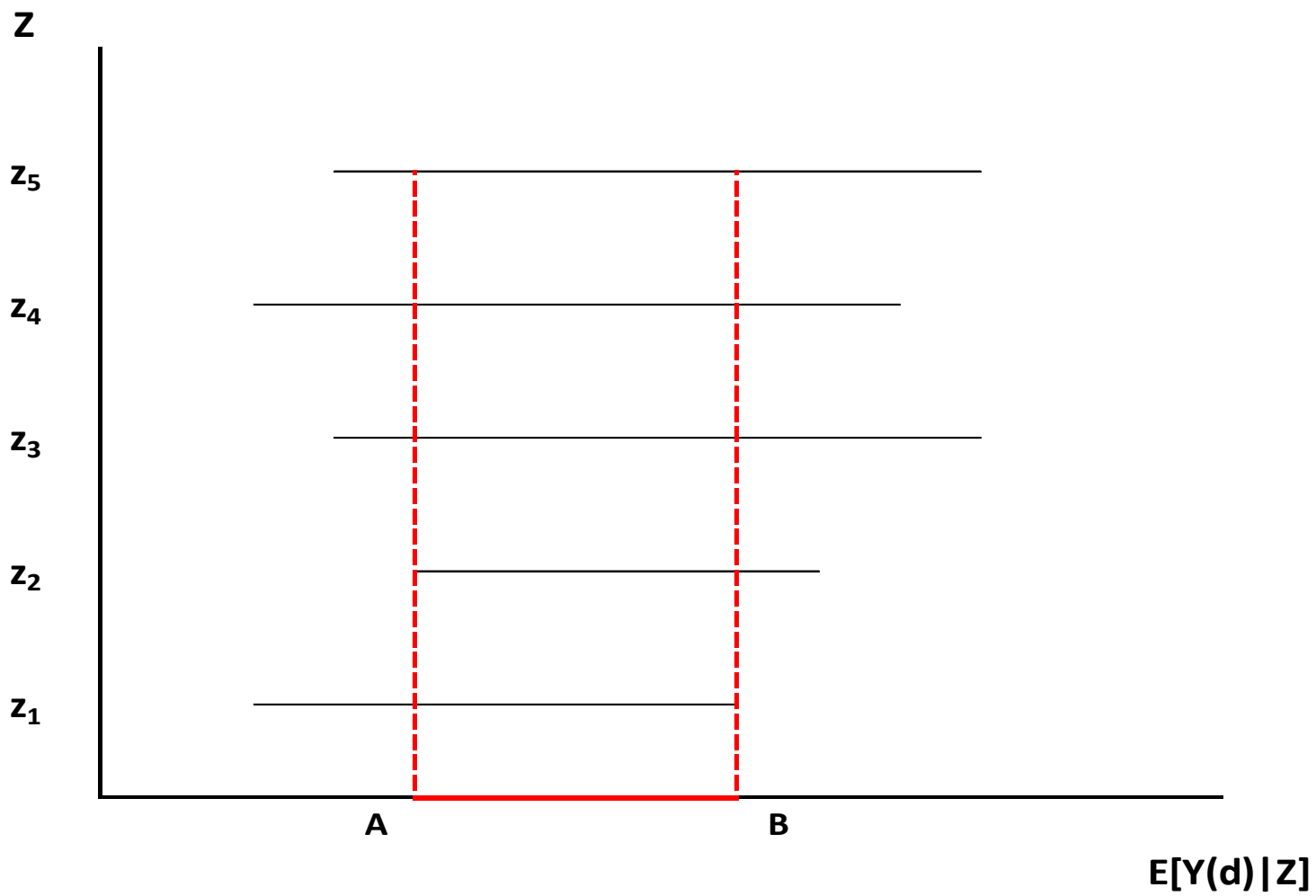


# How do MIVs work? - XIVs (cont.)

- Since the instrument is exogenous, results from all conditional identification regions are valid for the population
- All regions give valid answers → the intersection of all identification regions
- The intersection is given by the maximum lower bound and the minimum upper bound

$$\max_z LB[d|Z=z] \leq E[Y(d)] \leq \min_z UB[d|Z=m]$$

# How do MIVs work? - XIVs (cont.)

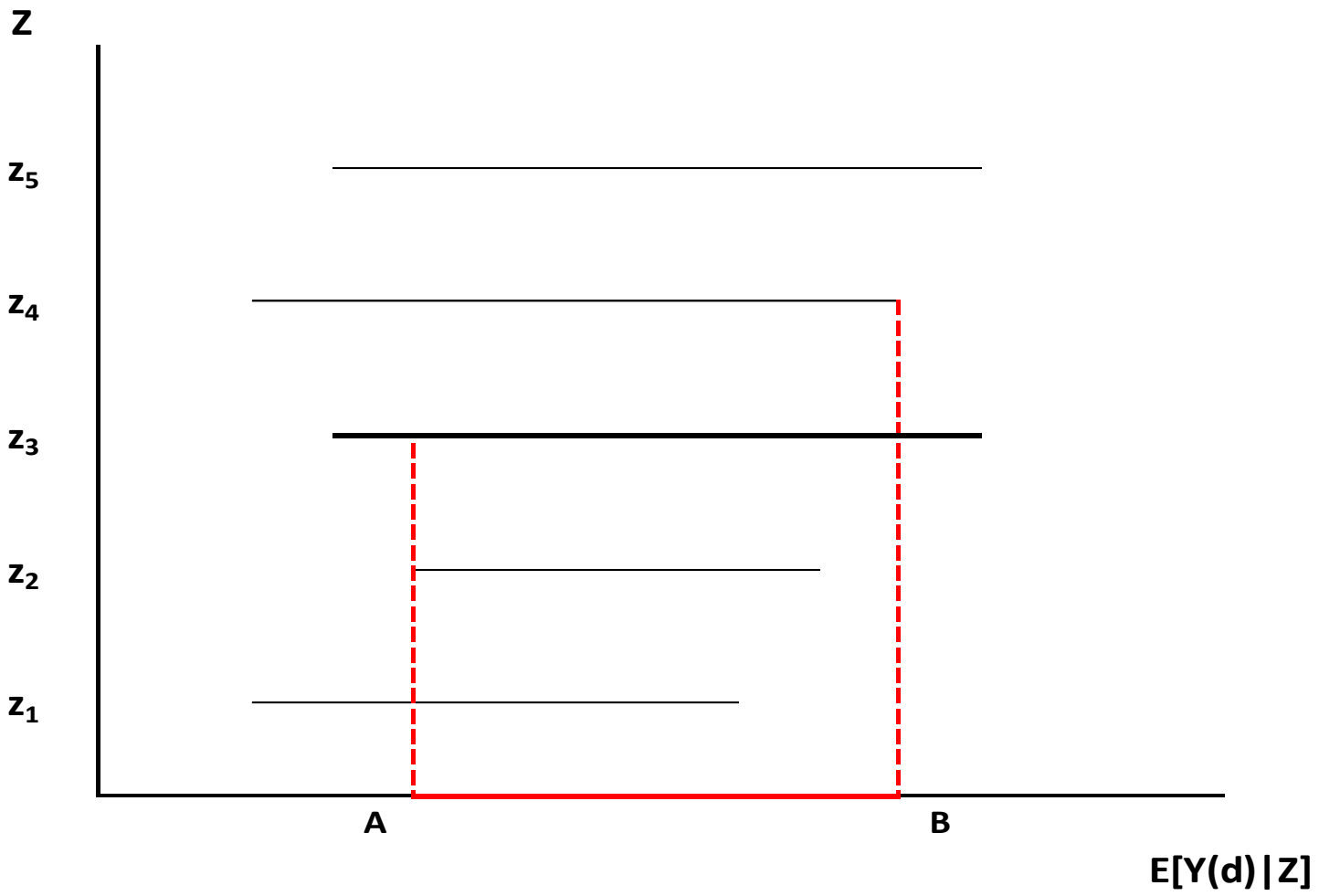


# How do MIVs work?

$$z_1 \leq z_2 \rightarrow E[Y(d)|Z = z_1] \leq E[Y(d)|Z = z_2]$$

- For a given MIV value can examine only what happens to
  - lower bounds at lower instrument values
  - upper bounds of higher instrument values

# How do MIVs work?



# How do MIVs work? (cont.)

- Hence, the final bounds for the identification region conditional on MIV value $z_3$ are:
  - the maximum lower bound over $z_1$, $z_2$, $z_3$, which is the one corresponding to MIV value $z_2$
  - the minimum upper bound over $z_3$, $z_4$, $z_5$, which is the one corresponding to MIV value $z_4$
- Analogous calculations performed for identification regions conditional on MIV values other than $z_3$
- Finally, overall bounds are computed by integrating out the MIV

# How do MIVs work? (cont.)

- Thus, optimization takes place only for a restricted range of values – identification regions less informative

- To be expected, as the MIV assumption is weaker than the XIV one

# MIVs (cont.)

- MIVs partition the sample space into cells defined by instrument values and look for minima (maxima) of upper (lower) bounds
- In the case of MIVs, the search is directed
- Typically, the finer the partition (achieved, e.g., by using more than one MIVs) the narrower the identification regions
- Local minima and maxima are perfectly OK – they just lead to wider identification regions
- Different instruments possibly lead to different identification regions – this is fine, as the unobserved average potential outcomes do not change
- Constraint: number of observations