# Treatment Effects With Unobserved Heterogeneity: A Set Identification Approach using an Instrumental Variable

Domenico Depalo *

**Abstract**

This paper introduces several sharp bounds of the potential outcome distributions for *panel data*. Jun et al. (2016) propose a design that allows for non-random treatment assignments until unobserved heterogeneity is controlled for. Using the same design, the bounds of this paper complement those in Jun et al. (2016) because are valid when more structure can be imposed or an instrumental variable (IV) exists. The gain from exploiting more information is shown by mean of two examples: one *ad hoc* and one on real patients-level data to study the relationship between adherence to medical prescription and reduction of cholesterol. Coherently with most recent results, the merit of drugs alone to reduce cholesterol is not so clear. The new bounds are much narrower than those available - not exploiting the longitudinal dimension of the data or not exploiting a IV.

JEL classification: C14, C23, C26, I12

Keywords: bounds, panel data, endogeneity

*I would like to thank Monica Andini, Vincenzo Atella, Yoonseok Lee. Replication files and additional results will be available at the webpage: http://sites.google.com/site/domdepalo/ The views expressed in this paper are those of the author and do not imply any responsibility of the Bank of Italy. Corresponding address: Domenico Depalo, Banca d'Italia, Economics and Statistics Department, Via Nazionale, 91 - 00184 Rome (Italy), Tel.: +39-06-4792 5989, e-mail: domenico.depalo@bancaditalia.it

# 1 Introduction

Set identification is increasingly popular for applied research. The methodological research initiated by Manski (1990) aims at relaxing strong, often untestable, hypotheses that allow point identification.

Early contributions to this literature are based on cross sectional setting. Bounds derive taking advantage of specific features of the data, but not the time dimension. With the availability of repeated observations for each single unit (i.e., panel data setting), exploiting the time dimensionality is a promising direction to obtain further information. To the best of my knowledge, only few solutions are available for this setup. Arpino et al. (2014) used a longitudinal survey to bound the prevalence of Human Immunodeficiency Virus (HIV) in rural Malawi, exploiting the absorbing nature of the infection. Jun et al. (2016) estimate the effect of smoking during pregnancy on infant birthweight, using a setup very similar to the standard Fixed Effect (FE) models. To bound the cumulative distribution function (CDF) of the birthweight, they take advantage of the admissible values of CDFs.

This paper uses a setting similar to Jun et al. (2016). Its main motivation is the investigation of other features of the data that allow to impose further structure. The benefit from doing this is that the bounds become narrower. The features investigated are common to a large class of empirical models and thus may be of large practical relevance. However, it should be born in mind that partial identification is not a panacea for using assumptions (Ho and Rosen (2015)).

Bounded outcome bounds are a useful starting point of each analysis. Yet, moving away from them is a necessary step to make the *panel setup* an appealing alternative to *cross sectional setup*. To this aim, I first investigate a well known assumption in the literature, namely the mean independence as in Manski (1990). This is a key step forward because several bounds for cross sectional setting take bounds under this hypothesis as a benchmark. Second, I propose narrower bounds that exploit other characteristics of the data, namely the first order stochastic dominance -maybe over a smaller support of the data- as proposed by Blundell et al. (2007).

I show how the methods work by mean of two examples: the first uses *ad hoc* data where all the

ideal conditions are met; the second uses real data from the Health Search Database (HSD), a longitudinal observational database collected by the Italian College of General Practitioners (SIMG). In the latter example, I estimate the relationship between adherence to medical prescription and reduction of cholesterol. In contrast to much of the existing literature, that highly emphasizes the role of drugs to reduce cholesterol, once the uncertainty about the unobserved data and the longitudinal dimensions are taken into account, the merit of drugs alone to reduce cholesterol becomes less clear, confirming recent findings in Atella et al. (2011).

Both applications show that exploiting the longitudinal dimension of the data indeed provides more information than approaches that do not exploit it. As either the time dimensionality of panel data and/or the share of observed units contributing to the Cumulative Distribution Function get larger, the width of panel data bounds becomes smaller. Exploiting the information provided by an external IV narrow the bounds even further.

The paper is organized as follows. In Section 2 I briefly review bounds for the distribution function of outcome variables in cross sectional settings (more details can be found in Appendix I): it should be clear at the outset that there is nothing conceptually new in these bounds. The review of cross sectional bounds is useful to set the ground for panel data bounds, presented in Section 3. In Section 4 I show the *ad hoc* example. As a more concrete example, in Section 5 I use data from an Italian sample of patients suffering from high cholesterol. Section 6 offers some conclusions.

## 2 Cross sectional bounds

The aim of the analysis is the identification of the Cumulative Distribution Function of a potential outcome variable $y$, exposed to a treatment $D \in \{0,1\}$. Throughout the paper, I assume that everything is conditional on adjusting variable(s) $X$, but I suppress it for a succinct presentation. In general, the CDF cannot be estimated by observed data. To see this, note that by definition $F^j(y) = F(Y^j \leq y|D = j)\,\mathbb{P}(D = j) + F(Y^j \leq y|D = 1-j)\,\mathbb{P}(D = 1-j)$. Whilst I observe $F(Y^j \leq y|D = j), \mathbb{P}(D = j)$ and $\mathbb{P}(D = 1-j) = 1 - \mathbb{P}(D = j)$, I do not observe $F(Y^j \leq y|D = 1-j)$. In a nutshell, what bounds do is exploiting all the features of the data that 'shed light' on

$F(Y^j \leq y|D = 1-j)$.

In this section I focus on cross sectional bounds that will be generalized to panel data setting in Section 3. A recent review on bounds can be found in Ho and Rosen (2015). The 'bounded outcome bounds' of Manski (1994) are such that: $\mathbb{P}(Y^j \leq y, D = j) \leq F^j(y) \leq \mathbb{P}(Y^j \leq y, D = j) + \mathbb{P}(D = 1-j)$. The 'independence bounds' of Manski (1990) & Heckman and Vytlacil (2000) take the form (focusing on $j = 1$): $\mathbb{P}(Y^j \leq y, D = j|Z = 1) \leq F^j(y) \leq \mathbb{P}(Y^j \leq y, D = j|Z = 1) + \mathbb{P}(D = 1-j|Z = 1)$. The 'First Order Stochastic Dominance (FSD) bounds' of Blundell et al. (2007) are equal to: $\mathbb{P}(Y^j \leq y|D = j) \leq F^j(y) \leq \mathbb{P}(Y^j \leq y|D = j)\,\mathbb{P}(D = j) + \mathbb{P}(D = 1-j)$. A 'weak form of FSD bounds' proposed by Blundell et al. (2007), for a given quantile Q and $y^{Q(D=1)}$ the correspondig Q-th quantile of $y$, are: if $y < y^{Q(D=1)}$ then $\mathbb{P}(Y^j \leq y|D = j)\,\mathbb{P}(D = j) \leq F^j(y) \leq \mathbb{P}(Y^j \leq y|D = j)\,\mathbb{P}(D = j) + \mathbb{P}(D = 1-j)$ and if $y \geq y^{Q(D=1)}$ then $\mathbb{P}(Y^j \leq y|D = j)\,\mathbb{P}(D = j) + Q\,\mathbb{P}(D = 1-j) \leq F^j(y) \leq \mathbb{P}(Y^j \leq y|D = j)\,\mathbb{P}(D = j) + \mathbb{P}(D = 1-j)$.

In Appendix I, I provide all the necessary details. In Table 1, I report relevant information regarding the assumptions and the bounds that follow.

# 3 Panel data bounds

When the data are available for the same unit at different points of time, the conditions of Section 2 should be refined to exploit the panel dimensionality. Then, it is possible to recover more information with respect to the cross sectional setting. Under the hypotheses outlined below, with panel data part of the information can be recovered from any of the repeated observations. Jun et al. (2016) propose sharp bounds of the cumulative distribution functions of potential outcomes in a situation where treatment assignments are not necessarily exogenous because of the presence of unobserved heterogeneity. In their Fixed Effect (FE) model presented later, they exploit only the admissible values of a CDF, coherently with the data and the problem at hand.

A contribution of the present paper is to make the bounds in Jun et al. (2016) narrower, possibly taking advantage of an external exogenous information if it is available. This is a simple, yet necessary, step forward. In the cross sectional setting, a large number of the extensions of

bounds are based on the existence of an exogenous variation. The independence bounds of Manski (1994) is the most prominent benchmark against which to compare other bounds. Therefore, having a counterpart for panel data is very useful for further developments.

Throughout the section I consider $\{(Y_{i,t}, D_{i,t})\} : i = 1, 2, \ldots, N; t = 1, 2, \ldots, T$. I also assume that all random variables are identically distributed across $i$. The only difference with respect to Section 2 is that all the variables are now indexed by individual and time dimensionalities. The notational difference is minimal, but the impact on identified bounds will be huge.

### Bounded outcome bounds of Jun et al. (2016)

The formal specification of the FE model of Jun et al. (2016) is based on two assumptions:

- *Selection on unobservable*: There exists a vector $\alpha_i$ of time-invariant unobserved confounding factors such that for $j = \{1, 0\}$ and for $\{t = 1, 2, \ldots, T\}$, $Y_{it}^j$ is independent of $D_{it}$ conditional on $\alpha_i$.

- *Time homogeneity*: For $j = \{1, 0\}$ and for $\{t = 1, 2, \ldots, T\}$, $Y_{it}^j$ and $Y_{i1}^j$ have the same distribution conditioning on $\alpha_i$.

An extensive explanation of what these assumptions imply is in Jun et al. (2016). Most important, no restriction is imposed on $\alpha_i$ which thus *1)* is equivalent to the individual fixed effects in panel-data models and *2)* can be arbitrarily correlated to the endogenous treatment $D_{i,t}$. Therefore, this model can accommodate situations where $D_{i,t}$ is potentially endogenous and the selection is on unobservables characteristics.

For the present paper, it is worth emphasizing that *selection on unobservable* concerns only the marginal outcome distributions; *time homogeneity* allows us to focus on $F^j(y)$ instead of on $F_t^j(y)$.

Under these hypotheses, by definition of CDF it is possible to impose the assumption (from now on, I drop the subindex $i$ from notation, unless necessary):

- *Bounded outcome-P*: $0 \leq \mathbb{P}(D_1 = D_2 = \ldots = D_s = 1 - j) \leq 1$.

Jun et al. (2016) derive sharp bounds of the CDf as

$$L^j(y) \leq F^j(y) \leq U^j(y) \tag{1}$$

where $L^j(y) = \sum_{s=1}^{T} p_s^j(y)$ with $p_1^j(y) = \mathbb{P}(Y_1 \leq y, D_1 = j)$ and $p_s^j(y) = \mathbb{P}(Y_s \leq y, D_1 = D_2 = \ldots = D_{s-1} = 1-j, D_s = j)$ for $s = 2, \ldots, T$ and $U^j(y) = L^j(y) + \mathbb{P}(D_1 = D_2 = \ldots = D_s = 1-j)$. Some comments regarding these bounds are in order (see also the original paper). First, the form of bounds clarifies that these hypotheses may not be innocuous. For example, -a bit heuristically- time homogeneity allows recovering the information regarding $F_1^j(y)$ from $F_2^j(y)$: if the two CDFs are different (i.e., if time homogeneity is violated), then it is no longer clear what I am bounding. Whether the hypothesis is credible or not should be checked case-by-case. When $T$ is small, time homogeneity may be more credible than when $T$ is large. More generally, in Section 5 I propose to use the Oaxaca (1973) and Blinder (1973) decomposition to support the time homogeneity, even when $T$ increases. Second, and most important, the width of bounds decreases as $T$ increases and point identification is achieved as $T \to \infty$, since $\mathbb{P}(D_1 = D_2 = \ldots = D_s = 1-j) \to 0$. Third, for subsamples for which $\mathbb{P}(D_1 = D_2 = \ldots = D_s = 1-j) = 0$, $F^j(y)$ is point identified. Fourth, unlike the local identification power of standard IV methods (LATE as proposed by Imbens and Angrist (1994); Angrist et al. (1996)), this subpopulation is identifiable from the data (a point first emphasized by Manski and Pepper (2000)). In the empirical application of Section 5 I show these properties at work.

### Independence bounds of Manski (1990) & Heckman and Vytlacil (2000)

The treatment indicator $D_{i,t}$ may be the result of a individual utility maximization. In a job training program only those individuals who feels to be successful may be induced to participate (as found for example by Lee (2009) using bounds to estimate the effect of the Job Corps program, one of the largest federally funded job training programs in the U.S.); the return on education may reflect the ability of the workers (for a comprehensive treatment of this issue see Card (1995, 1999)); the causal effect of catheterization on health outcomes cannot be estimated by comparing

outcomes between catheterized and non-catheterized patients, if the two groups differ on unobserved dimensions (as found by Bhattacharya et al. (2008, 2012)); related to Section 5, Egan and Philipson (2014) propose an economic model of patients' non-adherence to prescriptions of practitioners as an optimal stopping problem run by patients, based on their response to therapy.

Although controlling for $\alpha_i$ already removes much of the unobserved heterogeneity and endogeneity, part of them will not be eliminated (Vella (1998)). In the previous examples, it may be argued that $\alpha_i$ is enough for the estimation of return on education (if ability is time invariant). However, $\alpha_i$ does not fully capture all the features of the optimal stopping problem relevant to explain the patients' non-adherence, as part of these features are likely to be time varying (at least for the rational patient). In these situations, an exogenous instrumental variable that is independent of potential outcomes is needed.

Moreover, whilst in a classical FE setup having a $Z_{it} = Z_i$ would be problematic because $Z_i - \sum_1^T Z_{it}/T = Z_i - \sum_1^T Z_i/T = 0$, with these bounds it is possible to set $\alpha_i = (\alpha_i, Z_i)$. This definition of instrumental variable is more general than it may appear at first glance. For example, participation to several programs are based on a (form of) lottery, whose extraction is kept fixed over a large horizon, in which case $Z_{i1} = Z_{i2} = \ldots = Z_{iT} = Z_i$ (as in Lee (2009)); typical instruments to estimate the return on education are family background (Card (1999) for a review and a critique) and the quarter of birth (Angrist and Krueger (1991)), such that $Z_{i1} = Z_{i2} = \ldots = Z_{iT} = Z_i$; patients are excluded from drug payment if disabled or retired (plus other conditions), in which case $Z_{i1} = Z_{i2} = \ldots = Z_{iT} = Z_i$. Hence, the following bounds are obtained under hypotheses similar to those in Jun et al. (2016), conditioning on $(\alpha_i, Z_i)$ rather than $\alpha_i$ alone.[1] This accommodates situations where the selection is based on characteristics that are partially observable ($Z_{it}$) and partially not observable ($\alpha_i$). To be more explicit on the role of the instrument $Z_{it}$, I introduce the following assumptions:

- *Treatment assignment*: Statistical randomization of treatment assignment is not achieved until *unobserved* heterogeneity is controlled for and an appropriate (see last condition)

[1] In some bounds presented later, the conditioning on $Z_i$ is redundant. I decided to leave it anyway to emphasize its role, with respect to bounds in Jun et al. (2016) that do not need it.

*observable* instrument is used.

- *Time homogeneity of Z*: For $j = \{1, 0\}$ and for $\{t = 1, 2, \ldots, T\}$, $Z_{it}$ and $Z_{i1}$ have the same distribution conditioning on $\alpha_i$;

this allows using $Z$ instead of $Z_t$; the same condition applies to $Y_{it}$.

- *Independence of Z-P*: Conditional on $\alpha_i$, $Y_{it}^j$ are independent of Z.

Imposing these assumptions, there exists a Z such that $g(y_d|Z, \alpha_i) = g(y_d|\alpha_i)$ –suppressing the subindex $i$. Using Theorem 2 of Heckman and Vytlacil (2000), it immediately follows that

$$L^j(y) \leq F^j(y) \leq U^j(y) \tag{2}$$

where $L^j(y) = \sum_{s=1}^{T} p_s^j(y)$ with $p_1^j(y) = \mathbb{P}(Y_s \leq y, D_1 = j|z = j)$, $p_s^j(y) = \mathbb{P}(Y_s \leq y, D_1 = D_2 = \ldots = D_{s-1} = 1 - j, D_s = j|z = j)$ for $s = 2, \ldots, T$ and $U^j(y) = L^j(y) + \mathbb{P}(D_1 = D_2 = \ldots = D_s = 1 - j|z = j)$. These bounds retain the same properties of bounds from Jun et al. (2016). In particular, they are sharp. To see this, apply the proof of Theorem 1 by Jun et al. (2016), conditioning on Z. The conditioning leaves the proof identical because, by assumption *independence of Z-P*, it follows that $g(y_d|Z, \alpha_i) = g(y_d|\alpha_i)$. Therefore, the proof provided by Jun et al. (2016) is valid even under this new larger set of hypotheses.

### First Order Stochastic Dominance (FSD) bounds of Blundell et al. (2007)

The panel data bounds of Jun et al. (2016) can be narrowed if one is willing to impose the first order stochastic dominance (FSD) as in Blundell et al. (2007).

In the standard labor supply model, individuals with higher wages will be more likely to work unless the difference between wages and reservation wages is negatively associated with wages (that motivated the restriction in Blundell et al. (2007)). In the above examples on job training, schooling, catheterization, or drug prescription, it may be reasonable to assume that these treatments are beneficial for some individuals, but not detrimental for the rest of the population.

In these cases it is reasonable to impose the first order stochastic dominance defined for panel data, over the whole period, that:

- *First Order Stochastic Dominance-P*

$$\begin{aligned} \mathbb{P}(Y_t^j \leq y | D_1 = j) &\leq \mathbb{P}(Y_t^j \leq y | D_1 = 1-j) \qquad \& \\ \mathbb{P}(Y_t^j \leq y | D_1 = D_2 = \ldots = D_{s-1} = 1-j, D_s = j) &\leq \mathbb{P}(Y_t^j \leq y | D_1 = D_2 = \ldots = D_s = 1-j) \qquad \forall s. \end{aligned}$$

The bounds are narrower, as

$$L^j(y) \leq F^j(y) \leq U^j(y) \tag{3}$$

where $L^j(y) = \sum_{s=1}^{T} p_s^j(y)$ with $p_1^j(y) = \mathbb{P}(Y_1 \leq y | D_1 = j)$ and $p_s^j(y) = \mathbb{P}(Y_s \leq y | D_1 = D_2 = \ldots = D_{s-1} = 1-j, D_s = j)$ and $U^j(y)$ is defined in equation 1.

### Weakening FSD

Whether FSD is valid over the entire support of the outcome distribution or over a smaller subset may be a matter of disagreement. More generally, for some groups the stochastic dominance assumption may not be verified. Following Blundell et al. (2007), I propose the weaker restriction that the Q-th quantile of unobserved outcome is not higher than the Q-th quantile of the observed outcome, over the whole period. Hence:

- *Q-th quantile-P*

$$\begin{aligned} 0 &\leq \mathbb{P}(Y_t^j \leq y | D_1 = D_2 = \ldots = D_s = 1-j) \leq 1 \quad \text{if } y < y^{Q(D=1)} \qquad \forall s \\ Q &\leq \mathbb{P}(Y_t^j \leq y | D_1 = D_2 = \ldots = D_s = 1-j) \leq 1 \quad \text{if } y \geq y^{Q(D=1)} \qquad \forall s \end{aligned} .$$

In the example of Blundell et al. (2007) this assumption is equivalent to assume, for each time period, that the Q-th quantile wage offer for those not working is not higher than the Q-th quantile observed wage $y^{Q(D=1)}$. If $Q = 0.5$ the probability of observing someone working who has a wage above the median is higher than if the wage is below it. The same rationale applies to the estimates of return to education. In the stopping model proposed in Egan and Philipson (2014),

the probability of non-stopping the therapy when the improvement in health condition is above the Q-th quantile is higher than when the improvement is below this threshold.

Then, the bounds of Jun et al. (2016) become:

$$L^j(y) \leq F^j(y) \leq U^j(y) \tag{4}$$

where $L^j(y)$ and $U^j(y)$ are those from equation 1 if $y < y^{Q(D=1)}$, and $L^j(y) = \sum_{s=1}^{T} p_s^j(y) + Q\,\mathbb{P}(D_1 = D_2 = \ldots = D_s = 1 - j)$ and all the other components as defined in equation 1 if $y \geq y^{Q(D=1)}$.

If the FSD –maybe over a narrower support– restriction is valid conditioning on an IV, then the last two bounds hold conditioning on Z (i.e., the relevant components are defined like in equation 2, rather than in equation 1).

## 4 An illustrative example

In this section I show the virtue of the larger set of assumptions exploited in this paper by mean of an illustrative example. In Figure 1 I apply the bounds of Sections 2–3 to *ad hoc* data. I set the number of periods equal to 2 and the number of individuals equal to 500, for a total of 1000 observations. The example is based on the following model:

$$\begin{aligned} y_d &= r(D) - \epsilon_d \\ D^* &= s(Z) - v \\ D &= 1(D^* \geq 0) \\ Z &\perp (\epsilon_0, \epsilon_1, v). \end{aligned}$$

The instrument Z is randomly drawn from a uniform distribution and takes value 1 when the draw is above 0.5 (i.e., $E[Z = 1] = 0.5$). This is a valid exclusion restriction for the endogenous treatment indicator $D$. The strength of endogeneity goes from 0.3 (moderately low endogeneity), to 0.5 (moderate endogeneity) and to 0.8 (strong endogeneity). The individual specific intercepts are drawn from a uniform distribution multiplied by a factor of 10 (of course, this is just a rescaling

with no impact on results).[2]

In the left hand side of Figure 1 are plotted the cross sectional bounds, reviewed in Section 2: the bounds of Manski (1994) are the light area; the bounds of Manski (1990) and Heckman and Vytlacil (2000) under independence are the shadow area; the bounds of Blundell et al. (2007) under First Order Stochastic Dominance conditioning on Z are the dark area.

Cross sectional bounds do not exploit all the features of the data: although I observe twice the same individual, I do not take this information into account while estimating the distribution function. Some remarks are in order. First, while the bounded outcome bounds are sharp, they are very large. Second, a larger set of assumptions makes the bounds narrower: for instance, the bounds under the independence of Z are smaller than those using only the bounded outcome; correspondingly, bounds from Blundell et al. (2007) under FSD are smaller than bounds from Manski (1994) and Heckman and Vytlacil (2000) under independence.

On the right hand side of Figure 1 are plotted the bounds exploiting the panel dimensionality, proposed in Section 3. With respect to cross sectional bounds, I now exploit the fact that the same individual is observed twice. This has a huge impact on the width of the estimator which shrink by much for all proposed bounds even with only 2 periods. Within panel data bounds, as the set of assumptions increases, the width of bounds becomes smaller. These results hold for all the bounds.

This simple example is fruitful to show that *1)* panel data setting is helpful to shrink the bounds, because they exploit more information than the cross sectional setup and *2)* adding structure is powerful to narrow the width of bounds.

# 5 An empirical application

In this section I apply cross sectional and panel data bounds to real data from the Health Search Database (HSD), a longitudinal observational database collected by the Italian College of General Practitioners (SIMG). The sample is representative of the Italian adult population (Fabiani et al. (2004)).

[2]The same conclusions discussed in this section hold with 100 and 1000 individuals, that can be found on my website. On the website I also present the results when the degree of endogeneity is equal to 0 (no endogeneity).

The characteristics of the dataset have been extensively discussed in Atella et al. (2011). In this paper, for the year 2006-07, I consider male patients between 40 and 80 years old when they first entered the HSD, diagnosed of "pure hypercholesterolemia" (or familiar hypercholesterolemia) and with at least a prescription of statin medications.[3] After this selection, I am left with 124,350 observations. They correspond to 16,931 patients, most of which are observed in each of the eight quarters (about 80% or 13,542 patients). Including also patients observed for 7 quarters, the share of patients increases to about 85% (and to almost 90% including also those who are observed 6 periods): hence, in this paper I consider the attrition as completely at random (Little and Rubin (1986); Little (1995)) and focus only on the balanced sample. The treatment of attrition in a panel data setup is left for future research. A promising direction is Horowitz and Manski (2000).

The health indicator of interest is the ratio between serum levels of low-density lipoprotein (LDL) and total cholesterol, a better predictor of future heart attack risk than total cholesterol itself (Baigent et al. (2005)). The research question is the response of cholesterol to drugs. To this aim, the dependent variable is the quarterly difference of the ratio. A crucial aspect for policy makers is the extent to which patients follow the medical prescriptions, or compliance/(non-)adherence.

Although this paper employs data from Italy, that is based on a National Health System (NHS), non-adherence is a generalized problem around the world. According to estimates provided in a report by the New England Healthcare Institute (2009), the annual cost of non adherence in the United States is approximately $290 billion (2.3% in terms of GDP). The conclusion reached in the report is that improving compliance to medical prescriptions is key to improve health outcomes and lowering health care costs. In this section I test whether the CDF of the reduction of cholesterol of 'non-switcher patients' (i.e., patients who are always low or high compliers over the period of the analysis) stochastic dominates the CDF of 'switcher patients'.

Even though the purpose of this section is only illustrative, before proceeding it is important to emphasize the similarities between the Italian health system and other health systems around the world. The most prominent similarity is the co-payment (i.e., patients must contribute towards the

[3]Hypercholesterolemia is the presence of high levels of cholesterol in the blood. It is not a disease but a metabolic derangement that can contribute to many forms of disease, most notably coronary artery disease. The primary therapy to treat hypercholestorelemia is statin medications which limit the body's production of cholesterol.

cost of their medication and health care use) such that patients are sensitive to the cost of drugs (as estimated on the same pathology using different Italian data by Atella et al. (2006)). Due to these similarities across different health systems, it is not really surprising that Atella et al. (2005) and Piette et al. (2004) find similar evidence between Italy, UK and USA regarding the relation between the out-of-pocket drug expenditure and compliance for pathologies similar to those analyzed here.[4]

Following a large portion of the literature (WHO (1999); Vermeire et al. (2001); Chernew et al. (2008) and the literature therein), I proxy compliance with the Mean Possession Ratio that, for each patient, is obtained as the ratio between the number of prescribed days of treatment and the number of days between prescription refills. As long as this ratio is higher than 0.75, the patient is defined as highly adherent (this is the split followed in Atella et al. (2011), who also run some robustness check regarding this threshold). In the sample at hand, in about 60% of observations the compliance is classified as 'high'.

Whether an individual is compliant or not may be endogenously determined. Part of this endogeneity may be due to time invariant individual unobservable characteristics (recently emphasized by James et al. (2014)), like risk aversion of the patients; part of the compliance beahviour may be due to a cost/benefit analysis, where cost is the out-of-pocket expenditure for drugs and the benefit is the health status. Egan and Philipson (2014) interpret non-adherence as an optimal stopping problem for patients learning about their individual value of a therapy. Indeed, patients are more informed about their reaction to the treatment, taking into account treatment effectiveness, side effects and costs of care.

This example fits perfectly in the setup of this paper. Unobservable time invariant individual characteristics will be absorbed in the $\alpha_i$s, whereas the cost/benefit channel may be better explained using an instrumental variable that affects the compliance but not the cholesterol reduction. To this aim, I exploit an indicator of expenditure exemption. In the Italian NHS, exemption is due to several reasons: wage, age, pathology, disability and a residual category. This instrument has no impact on the quarterly difference of cholesterol (i.e., it is exogenous), but it is relevant to explain individual consumption of drugs and ultimately the compliance. As standard in IV setup, a reason

[4] Although obvious, an important caveat when interpreting the results is that it is unclear if and how the results of this paper may be generalized to other countries (and I will not try to generalize the results of this paper).

of concern is related to the legitimacy of this instrument with respect to exogeneity. Undoubtedly, exemption motivations might have an impact on the *level* of cholesterol, but they hardly affect its *evolution over time*. To further mitigate any concern, I lagged the indicator of exemption a year earlier. To check whether these arguments are tenable, I run three regressions: apart from adjusting covariates, one of $y_t$ on $f(z_{t-4})$, one of $y_{t-1}$ on $g(z_{t-4})$ and one of $\Delta y_t$ on $\delta(z_{t-4})$. I successfully tested that $f(z_{t-4}) = g(z_{t-4})$ and that $\delta(z_{t-4}) = 0$. These results are only suggestive that the instrument works. However, rather than proceeding 'as if' the instrument is valid, I formally tested the assumptions of independence and monotonicity as suggested in Mourifié and Wan (2016), that -due to monotonicity- are more restrictive than those needed here. I did not reject the null hypothesis: the validity of the instrument is not falsified.

Slightly less than half of the patients are exempted from drug expenditure (5,313 patient, or about 45%).[5] The correspondence between expenditure exemption and high compliance is remarkable: about 31% of exempted patients are high compliers; the lowest share (less than 20%) are exempted from drug payment but are not high compliers (they are likely to be never-takers in the terminology of Angrist et al. (1996) and Imbens and Angrist (1994)); at the opposite sight, the share of those who are not exempted but still are high compliers is about 25% (they are likely to be always-takers).

The first order stochastic dominance assumption would be satisfied if drugs are not harmful (e.g., no patient gains from drugs, apart from at least only one who actually enjoy a benfit). The selection of patients described in Atella et al. (2011) guarantees that these individuals do not suffer from other pathologies, thus the assumptions seems trustworthy and is maintained throughout the rest of the analysis.

A critical assumption for these panel data bounds is time homogeneity on the potential outcomes. I test if it is credible or not by mean of the Oaxaca (1973) and Blinder (1973) decomposition on the outcome. If the coefficients are different as going from one time period to the next, this would be evidence of violation of time homogeneity. Yet, it is important to understand that non rejecting the equality of coefficients over time does not imply that the hypothesis is true, e.g. be-

[5]For 1,472 patients it is unknown whether they are exempted or not. They are likely to pay, without providing any documentation.

cause the test depends on the model specification. Apart from the exception in 2006:Q2, over the 8 quarters of the analysis I never reject that coefficients are equal over time. With a time period of only 2 years (i.e., 8 quarters), I interpret this as evidence that time homogeneity assumption might not be too strong in this application.

With this background it is interesting to test the effect of compliance on cholesterol reduction focusing only on patients who never change their compliance category ('No switch') as opposed to those whose adherence changes at some point ('Only switch'). To better emphasize the gain of the larger set of assumption in terms of identification, I do not show the confidence interval (see Horowitz and Manski (2000) for a similar argument). The method in Imbens and Manski (2004) can be implemented to this aim.[6]

In Figure 2 I plot bounds calculated according to Sections 2–3. Some comments are in order.

1. Using cross sectional bounds, there in no much difference between the two samples: this is because I cannot take advantage of the longitudinal dimension of the data and of Assumptions in Jun et al. (2016).

2. When I focus on panel bounds, the width for the sample of switchers is always smaller than for non-switchers, because for patients who are always high (low) compliers $\mathbb{P}(D_1 = D_2 = \ldots = D_s = 1 - j)$ remains constant by construction; in contrast, for switchers, it gets closer to zero as the share of switchers increases.

3. As a consequence of comment 2, only for switcher the width shrink by much with number of periods.

   This set of preliminary comments are the empirical counterparts to the remarks of bounds of Jun et al. (2016) in Section 3.

4. Imposing more assumptions is very successful to reduce the width of the bounds (this is a strong motivation in favour of the bounds proposed in this paper).

[6] A STATA code based on this method is available upon request.

5. The properties of bounds in Jun et al. (2016) (see previous comments 1–3) apply to the proposed new bounds, obtained under a larger set of assumptions (as noticed in Section 3).

6. If anywhere, evidence of stochastic dominance is possible only when T=8, using bounds from Blundell et al. (2007): even in this case, the large uncertainty about the unobserved outcome is such that the distribution of switchers is contained in the distribution of non switchers, therefore I cannot draw firm conclusion in favour or against the property.

7. As a consequence of comment 6, a formal test would not reject the null hypothesis of equality of the distributions (against the first order stochastic dominance) and adding periods in the hope of finding stochastic dominance is a practice doomed to fail (because the CDF for non switchers would still remain large).

The policy content of this analysis for the Italian case is substantial. It is unclear whether the improvement of compliance to medical prescription alone substantially (i.e., in first order stochastic dominance metric) improves health outcomes, with respect to those patients who do not switch their compliance behaviour. This conclusion is coherent with results in Atella et al. (2011) who emphasize the role of individual characteristics (beside the role of practitioners), using data as those used in this application.

# 6 Conclusions

In this paper I presented a variety of bounds that can be used in a *panel data* setup. The usefulness of exploiting this setting is shown by contrasting *panel data* bounds to the corresponding *cross sectional* bounds, that are first reviewed.

Using the same framework in Jun et al. (2016), I make use of an instrumental variable. Under the conditions of this paper -that are intended to cover a fairly large array of situations- the proposed bounds are smaller than those in Jun et al. (2016). However, there is no superiority of bounds presented here than those in Jun et al. (2016): the two approaches, rather than being substitutes, are complementary. To the extent that the structure proposed in this paper is tenable

or an IV satisfying all the required conditions exists, there is a gain from the proposed new bounds; otherwise, one should refrain from using untenable assumptions (Ho and Rosen (2015)).

The gain from using panel data bounds is shown by mean of two examples: one based on *ad hoc* data that satisfy all the necessary assumptions and one based on real data from a longitudinal observational database collected by the Italian College of General Practitioners, to study the relationship between adherence to medical prescription and reduction of cholesterol. Coherently with most recent results (see Atella et al. (2011) for the same dataset), once the uncertainty about the unobserved data and the longitudinal dimensions are taken into account, the merit of more drugs alone to reduce cholesterol with respect to patients who do not switch their compliance behaviour becomes less clear. A comparison with approaches that do not exploit the longitudinal dimension of the data shows that this characteristic improves the performance of bounds by much.

## References


Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables (Disc: p456-472). *Journal of the American Statistical Association 91*, 444–455.

Angrist, J. D. and A. B. Krueger (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics 106*(4), 979–1014.

Arpino, B., E. De Cao, and F. Peracchi (2014). Using panel data for partial identification of human immunodeficiency virus prevalence when infection status is missing not at random. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 177*(3), 587–606.

Atella, V., F. Belotti, and D. Depalo (2011, January). Disentangling true determinants of drug therapy effectiveness: A double fixed effects approach. CEIS Research Paper 186, Tor Vergata University, CEIS.

Atella, V., P. R. Noyce, E. Schafheutle, and K. Hassell (2005). Affordability of Medicines and Patients' Cost Reduction Behaviors: Empirical Evidence Based on SUR Estimates from Italy and the United Kingdom. *Appl Health Econ Health Policy 4*(1), 23–35.

Atella, V., F. Peracchi, D. Depalo, and C. Rossetti (2006). Drug compliance, co-payment and health outcomes: evidence from a panel of Italian patients. *Health Economics 15*(9), 875–892.

Baigent, C., A. Keech, P. Kearney, and L. Blackwell (2005). Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet 366*, 1267–1278.

Bhattacharya, J., A. M. Shaikh, and E. Vytlacil (2008, May). Treatment Effect Bounds under Monotonicity Assumptions: An Application to Swan-Ganz Catheterization. *American Economic Review 98*(2), 351–56.

Bhattacharya, J., A. M. Shaikh, and E. Vytlacil (2012). Treatment effect bounds: An application to SwanGanz catheterization. *Journal of Econometrics 168*(2), 223–243.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources 8*(4), 436–455.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007, 03). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica 75*(2), 323–363.

Card, D. (1995). Earnings, schooling, and ability revisited. In S. Polachek (Ed.), *Research in Labor Economics*, Volume 14, pp. 23 – 48. JAI Press.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 30, pp. 1801–1863. Elsevier.

Chernew, M. E., M. R. Shah, A. Wegh, S. N. Rosenberg, I. A. Juster, A. B. Rosen, M. C. Sokol, K. Yu-Isenberg, and A. M. Fendrick (2008). Impact of decreasing copayments on medication adherence within a disease management environment. *Health Affairs 27*(1), 103–112.

Chernozhukov, V., I. Fernndez-Val, J. Hahn, and W. Newey (2013). Average and quantile effects in nonseparable panel models. *Econometrica 81*(2), 535–580.

Egan, M. and T. J. Philipson (2014, July). Health care adherence and personalized medicine. Working Paper 20330, National Bureau of Economic Research.

Fabiani, L., M. Scatigna, K. Panopoulou, A. Sabatini, E. Sessa, F. Donato, M. Marchi, R. Nardi, C. Niccolai, F. Samani, and G. Ventriglia (2004). Health search: istituto di ricerca della societá italiana di medicina generale: la realizzazione di un database per la ricerca in medicina generale. *Epidemiol and Prev 28*, 156–162.

Heckman, J. J. and E. J. Vytlacil (2000, August). Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect. NBER Technical Working Papers 0259, National Bureau of Economic Research, Inc.

Ho, K. and A. M. Rosen (2015, October). Partial identification in applied research: Benefits and challenges. Working Paper 21641, National Bureau of Economic Research.

Horowitz, J. and C. F. Manski (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association 95*(449), 77–88.

Imbens, G. W. and J. D. Angrist (1994, March). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–75.

Imbens, G. W. and C. F. Manski (2004, November). Confidence Intervals for Partially Identified Parameters. *Econometrica 72*(6), 1845–1857.

James, P. A., S. Oparil, and B. L. Carter (2014). 2014 evidence-based guideline for the management of high blood pressure in adults: Report from the panel members appointed to the eighth joint national committee (jnc 8). *JAMA 311*(5), 507–520.

Jun, S. J., Y. Lee, and Y. Shin (2016). Treatment effects with unobserved heterogeneity: A set identification approach. *Journal of Business & Economic Statistics 34*(2), 302–311.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies 76*(3), 1071–1102.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association 90*(431), 1112–1121.

Little, R. J. A. and D. B. Rubin (1986). *Statistical Analysis with Missing Data*. New York, NY, USA: John Wiley & Sons, Inc.

Manski, C. F. (1990, May). Nonparametric Bounds on Treatment Effects. *American Economic Review 80*(2), 319–23.

Manski, C. F. (1994). The selection problem. In C. Sims (Ed.), *Advances in Econometrics: Sixth World Congress*. Cambridge University Press.

Manski, C. F. and J. V. Pepper (2000, July). Monotone Instrumental Variables, with an Application to the Returns to Schooling. *Econometrica 68*(4), 997–1012.

Mourifié, I. and Y. Wan (2016). Testing local average treatment effect assumptions. *Review of Economics and Statistics na*, n/a–n/a.

New England Healthcare Institute (2009). Thinking outside the pillbox a system-wide approach to improving patient medication adherence for chronic disease.

Oaxaca, R. (1973, October). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review 14*(3), 693–709.

Piette, J. D., M. Heisler, and T. H. Wagner (2004). Cost-Related Medication Underuse- Do Patients With Chronic Illnesses Tell Their Doctors? *ARCH INTERN MED 164*, 1749–1755.

Shea, J. (1997). Instrument relevance in multivariate linear models: A simple measure. *The Review of Economics and Statistics 79*(2), 348–352.

Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources 33*(1), 127–169.

Vermeire, E., H. Hearnshaw, P. Van Royen, and J. Denekens (2001). Patient adherence to treatment: three decades of research. a comprehensive review. *Journal of Clinical Pharmacy and Therapeutics 26*(5), 331–342.

WHO (1999). *1999 World Health Organization*. WHO-Geneva.

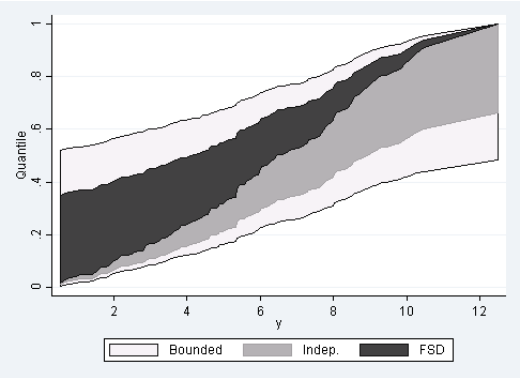
Table 1: Cross sectional bounds

| Method | Reference | Assumption | | Bounds |
|---|---|---|---|---|
| Bounded Outcome | Manski (1994) | $F(Y^j \leq y \vert D = 1-j)$ | L | $\mathbb{P}(Y^j \leq y, D = j)$ |
| | | $\in [0,1]$ | U | $\mathbb{P}(Y^j \leq y, D = j) + \mathbb{P}(D = 1-j)$ |
| Z-indep. | Manski (1990) | $D = 1[\mu(Z) \geq U]$ | L | $\mathbb{P}(Y^j \leq y, D = j \vert Z = 1)$ |
| | HV,2000 | $Z \perp (U, y_0, y_1)$ | U | $\mathbb{P}(Y^j \leq y, D = j \vert Z = 1) + \mathbb{P}(D = 1-j \vert Z = 1)$ |
| FSD | BGIM, 2007 | $\mathbb{P}(Y^j \leq y \vert D = j) \leq$ | L | $\mathbb{P}(Y^j \leq y \vert D = j)$ |
| | | $\mathbb{P}(Y^j \leq y \vert D = 1-j)$ | U | $\mathbb{P}(Y^j \leq y \vert D = j)\,\mathbb{P}(D = j) + \mathbb{P}(D = 1-j)$ |
| Q-th | BGIM, 2007 | $\mathbb{P}(Y^j \leq y \vert D = 1-j)$ | L | $\mathbb{P}(Y^j \leq y \vert D = j)\,\mathbb{P}(D = j)$ |
| | | $\in [0,1]$ if $y < y^{Q(D=1)}$ | U | $\mathbb{P}(Y^j \leq y \vert D = j)\,\mathbb{P}(D = j) + \mathbb{P}(D = 1-j)$ |
| | | $\mathbb{P}(Y^j \leq y \vert D = 1-j)$ | L | $\mathbb{P}(Y^j \leq y \vert D = j)\,\mathbb{P}(D = j) + Q\,\mathbb{P}(D = 1-j)$ |
| | | $\in [Q,1]$ if $y \geq y^{Q(D=1)}$ | U | $\mathbb{P}(Y^j \leq y \vert D = j)\,\mathbb{P}(D = j) + \mathbb{P}(D = 1-j)$ |

HV,2000: Heckman and Vytlacil (2000); BGIM, 2007:Blundell et al. (2007); L=lower bound; U=upper bound

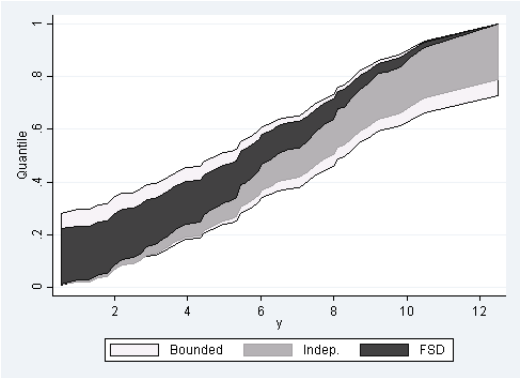
Figure 1: An example under ideal conditions met

Cross

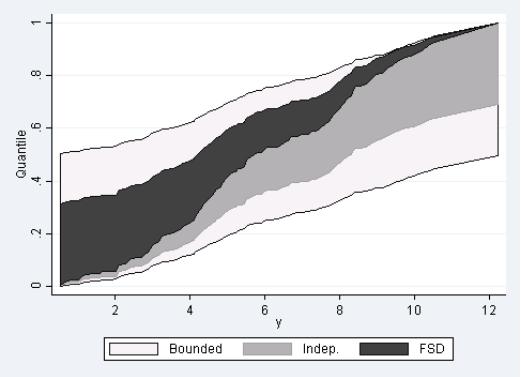



(a) Obs.:100 & $\rho = 0.3$

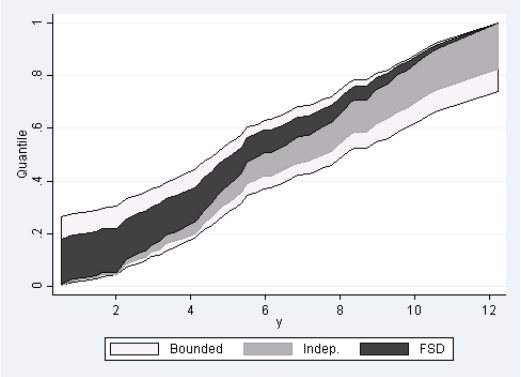
Panel



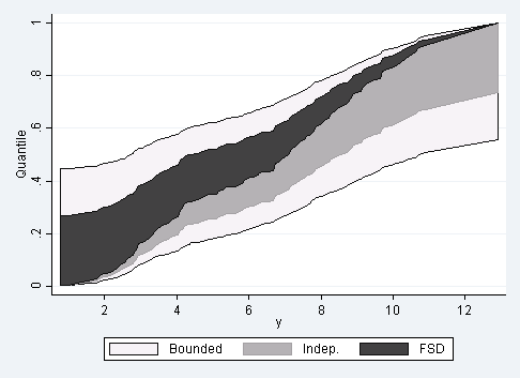

(b) Obs.:100 & $\rho = 0.3$



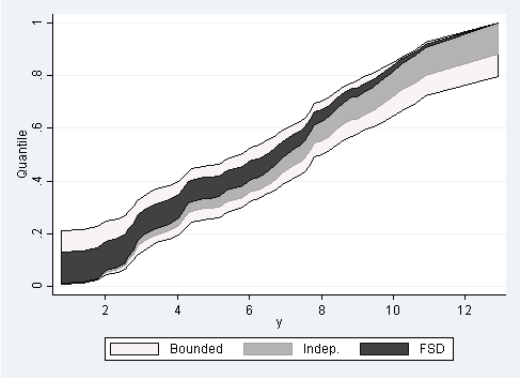

(c) Obs.:100 & $\rho = 0.5$




(d) Obs.:100 & $\rho = 0.5$



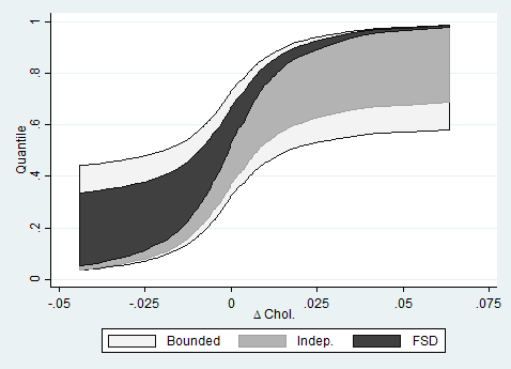

(e) Obs.:100 & $\rho = 0.8$



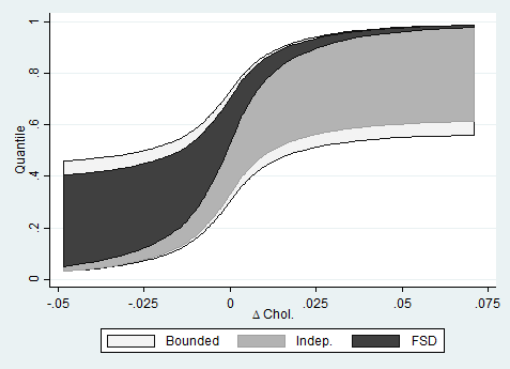

(f) Obs.:100 & $\rho = 0.8$

Figure 2: Compliance and cholesterol

Cross sectional bounds

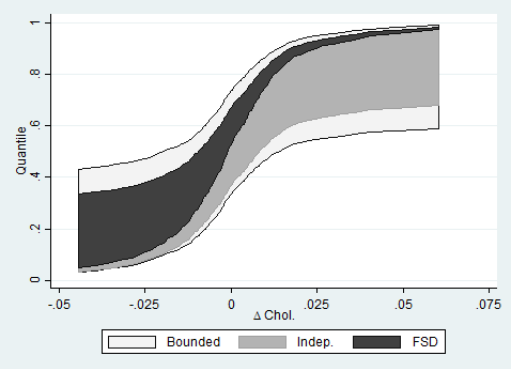



(a) No switch

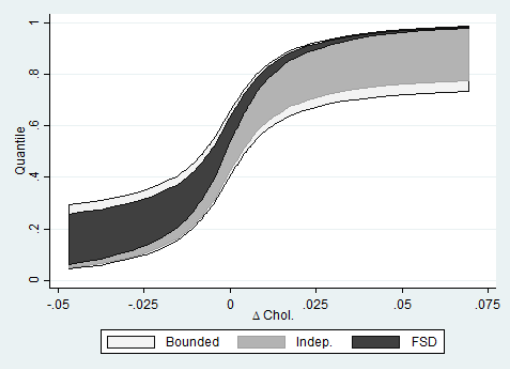



(b) Only switch

Panel bounds, T=4

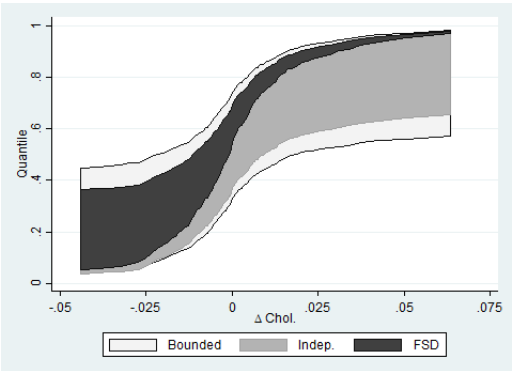



(c) No switch

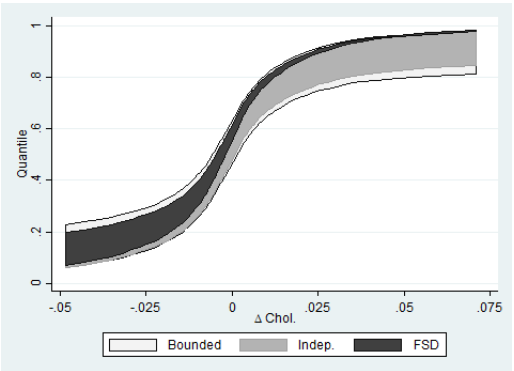



(d) Only switch

Panel bounds, T=8




(e) No switch




(f) Only switch

# I Details on cross sectional bounds

In this appendix I provide more details on cross sectional bounds discussed in Section 2 and shown in Table 1. The exposition for cross sectional bounds starts from the smallest set of assumptions, which imposes only that $F^j(y) \in [0,1]$, and then adds hypotheses, imposing the existence of an independent instrument and/or first order stochastic dominance. These bounds are not new, but are presented to make simpler the exposition of panel data bounds.

### Bounded outcome bounds of Manski (1994)

The first bounds that I present are from Manski (1994). They are based exclusively on the admissible values of a CDF, $F^j(y) \in [0,1]$. Therefore they are always valid, though usually quite large.

With these first bounds I provide all the details in an attempt to be as clear as possible. I am interested in $F^j(y)$. By definition, $F^j(y) = F(Y^j \leq y|D = j)\,\mathbb{P}(D = j) + F(Y^j \leq y|D = 1-j)\,\mathbb{P}(D = 1-j)$. The data are informative on $F(Y^j \leq y|D = j)$, $\mathbb{P}(D = j)$ and $\mathbb{P}(D = 1-j)$. Although I do not observe $F(Y^j \leq y|D = 1-j)$, it is possible to bound it between $\mathbb{P}_L(Y^j \leq y|D = 1-j) \leq \mathbb{P}(Y^j \leq y|D = 1-j) \leq \mathbb{P}_U(Y^j \leq y|D = 1-j)$, where $P_L(\cdot)$ and $P_U(\cdot)$ stand for lower and upper bounds, respectively. By definition of the CDF, it is possible to impose the assumption:

- *Bounded outcome*: $F(Y^j \leq y|D = 1-j) \in [k_0, k_1]$, with $\{k_0, k_1\} = \{0, 1\}$.

Taking advantage of this assumption, it follows that:

$$\begin{array}{ll}
\mathbb{P}(Y^j \leq y|D=j)\,\mathbb{P}(D=j) + \mathbb{P}_L(Y^j \leq y|D=1-j)\,\mathbb{P}(D=1-j) & \leq F^j(y) \leq \\
 & \mathbb{P}(Y^j \leq y|D=j)\,\mathbb{P}(D=j) + \mathbb{P}_U(Y^j \leq y|D=1-j) \\
 & \\
\mathbb{P}(Y^j \leq y|D=j)\,\mathbb{P}(D=j) + k_0\,\mathbb{P}(D=1-j) & \leq F^j(y) \leq \\
 & \mathbb{P}(Y^j \leq y|D=j)\,\mathbb{P}(D=j) + k_1\,\mathbb{P}(D=1-j) \\
 & \\
\mathbb{P}(Y^j \leq y|D=j)\,\mathbb{P}(D=j) + 0\,\mathbb{P}(D=1-j) & \leq F^j(y) \leq \\
 & \mathbb{P}(Y^j \leq y|D=j)\,\mathbb{P}(D=j) + 1\,\mathbb{P}(D=1-j) \\
 & \\
\mathbb{P}(Y^j \leq y, D=j) & \leq F^j(y) \leq \\
 & \mathbb{P}(Y^j \leq y, D=j) + \mathbb{P}(D=1-j).
\end{array} \tag{5}$$

These bounds become narrower as $\mathbb{P}(D=j)$ becomes larger, because as $\mathbb{P}(D=1-j) = 1-\mathbb{P}(D=j) \to 0$, it follows that $\mathbb{P}(Y^j \leq y, D=j) \leq F^j(y) \leq \mathbb{P}(Y^j \leq y, D=j) + 0$. This property holds throughout all the bounds that are presented for the cross sectional case.

### Independence bounds of Manski (1990) & Heckman and Vytlacil (2000)

The binary treatment $D$ may be endogenous (see Section 3 for some examples). In this case, the estimation requires an instrument $Z$ that influences the treatment indicator (i.e., relevant) but not the outcome (i.e., independence). More formally, following Heckman and Vytlacil (2000) I impose:

- *Z-indep.*: $D = 1\,[\mu(Z) \geq U]$ with $Z \perp (U, y_0, y_1)$.

This implies $g(y_d|Z) = g(y_d)$. This assumption is stronger than the simpler IV assumption that $E[y_d|Z] = E[y_d]$ because it implies a independence also between $Z$ and $U$ and because, in contrast to the IV assumption, is not just a *mean* independence assumption. Under *Z-indep.*

$$\begin{aligned}
\mathbb{P}(Y^j \leq y) &= \mathbb{P}(Y^j \leq y|Z=z) \\
&= \underbrace{\mathbb{P}(Y^j \leq y, D=j|Z=z)}_{observed} + \underbrace{\mathbb{P}(Y^j \leq y, D=1-j|Z=z)}_{NOTobserved}.
\end{aligned}$$

The last component is not observed, but by definition of CDF it is known that $0 \leq \mathbb{P}(Y^j \leq y | D = 1-j, Z = z) \leq 1$. By the same token of bounds of Manski (1994) :

$$\begin{array}{ll} \mathbb{P}(Y^j \leq y, D = j | Z = z) & \leq F^j(y) \leq \\ & \mathbb{P}(Y^j \leq y, D = j | Z = z) + \mathbb{P}(D = 1-j | Z = z) \\ max_z\{\mathbb{P}(Y^j \leq y, D = j | Z = z)\} & \leq F^j(y) \leq \quad \min_z\{\mathbb{P}(Y^j \leq y, D = j | Z = z) + \mathbb{P}(D = 1-j | Z = z)\}. \end{array} \tag{6}$$

Using the assumption *Z-indep.* and results in Heckman and Vytlacil (2000) (namely, equation 3 and Theorem 2), if $j = 1$ then

$$\mathbb{P}(Y^j \leq y, D = j | Z = 1) \leq F^j(y) \leq \mathbb{P}(Y^j \leq y, D = j | Z = 1) + \mathbb{P}(D = 1-j | Z = 1).$$

### First Order Stochastic Dominance (FSD) bounds of Blundell et al. (2007)

Bounds proposed by Blundell et al. (2007) are a natural extension of the bounds presented in Manski (1994), under bounded outcome, and Manski (1990) under independence. These bounds are valid whenever a FSD between treatment and non-treatment outcomes may be imposed to the data (Section 3 provides practical examples). The formal representation of this argument may be expressed as:

- *FSD*: $\mathbb{P}(Y^j \leq y | D = j) \leq \mathbb{P}(Y^j \leq y | D = 1-j)$.

Imposing this constraint, the lower bound is higher than in bounds of Manski (1994):

$$\begin{array}{ll} \mathbb{P}(Y^j \leq y | D = j) & \leq F^j(y) \leq \\ & \mathbb{P}(Y^j \leq y | D = j)\,\mathbb{P}(D = j) + \mathbb{P}(D = 1-j)\ . \end{array} \tag{7}$$

The difference in the lower bound with respect to equation 5 amounts to $\mathbb{P}(Y^j \leq y | D = j)\,\mathbb{P}(D = 1-j)$, where the first component is due to the assumption and the second component to the the population to which it applies.

If a valid IV exists, the same bounds hold conditioning on Z (Blundell et al. (2007)).

### Weakening FSD

If stochastic dominance is too strong for a subpopulation, it is possible to invoke a weaker restriction, valid from a quantile Q-th onward:

- *Q-th quant.*:

$$\begin{aligned} 0 &\leq \mathbb{P}(Y^j \leq y | D = 1-j) \leq 1 \quad \text{if } y < y^{Q(D=1)} \\ Q &\leq \mathbb{P}(Y^j \leq y | D = 1-j) \leq 1 \quad \text{if } y \geq y^{Q(D=1)}. \end{aligned}$$

The bounds that follow are

$$\begin{aligned} &\text{if } y < y^{Q(D=1)} : \\ &\mathbb{P}(Y^j \leq y | D = j)\,\mathbb{P}(D = j) && \leq F^j(y) \leq \\ & && \mathbb{P}(Y^j \leq y | D = j)\,\mathbb{P}(D = j) + \mathbb{P}(D = 1-j) \ ; \\ &\text{if } y \geq y^{Q(D=1)} : \\ &\mathbb{P}(Y^j \leq y | D = j)\,\mathbb{P}(D = j) + Q\,\mathbb{P}(D = 1-j) && \leq F^j(y) \leq \\ & && \mathbb{P}(Y^j \leq y | D = j)\,\mathbb{P}(D = j) + \mathbb{P}(D = 1-j) \ . \end{aligned} \tag{8}$$

These bounds are tighter than those from Manski (1994), as a function of $Q$, i.e. $[Q - \mathbb{P}(Y^j \leq y|D = j)]\,\mathbb{P}(D = 1-j)$ if $y \geq y^{Q(D=1)}$ and 0 otherwise, which also makes clear that the restriction provides tighter bounds to every quantile from the Q-th onwards.

If a valid IV exists, the same bounds are valid condition on Z (Blundell et al. (2007)).