

## A FORMAL TREATMENT OF IDENTIFICATION

ARIEL BOYARSKY\*

## 1. INTRODUCTION

Statistical inference is the process of extrapolating a sample to say something about the population. Identification on the other hand is what lets us go from the population to the parameters. More intuitively, it says that if we knew the population could we also learn the parameters. It is often seen as linking theory and data because identification often stands that we take a stance on the way the world works - in other words, we must make assumptions.

## 2. THE DEFINITION OF IDENTIFICATION

Suppose we observe some random vector,  $Y \in \mathbb{R}^{d_Y}$ , that is distributed  $G$ . We have that  $G \in \mathcal{G}$ , a set of possible distributions on  $\mathbb{R}^{d_Y}$ .

Now we want to recover some parameters of interest that describe  $G$ . We denote these parameters as  $\theta$ . Notice, that different parameters imply different distributions,  $G$ ,

$$\theta \mapsto G_\theta$$

. In other words, we can think that identification implies the existence of a mapping,  $\phi : \mathcal{G} \rightarrow \mathbb{R}^{d_\theta}$ . Where  $d_\theta$  is the dimension of the parameter. That is, we have some mapping that takes us from the population distribution to the parameter space – having such a mapping is exactly equivalent to identification.

We can now give a formal definition of identification via the set of parameters that can generate  $G$ ,

**DEFINITION 2.1 (The Identified Set)** *We say that a parameter belongs to the identified set if,  $\theta \in \Theta$  and  $G_\theta = G$ . Then we can characterize the identified set as,*

$$\Theta^*(G) \equiv \{\theta \in \Theta : G_\theta = G\}$$

The first condition,  $\theta \in \Theta$  implies that the parameter obeys any assumptions that we place on the parameter space. The second condition,  $G_\theta = G$  is observational equivalence. It says that we cannot tell apart the distribution of  $Y$  from the distribution that is implied by  $\theta \in \Theta^*(G)$ . This yields three possible cases,

1. Point Identification:  $|\Theta^*(G)| = 1$
2. Partial Identification:  $|\Theta^*(G)| > 1$
3. Misspecification:  $|\Theta^*(G)| = 0$

Notice that these variations of the identified set imply that the mapping,  $\phi(\cdot)$ , need not be well defined. That is, it could map us from a particular distribution to multiple  $\theta$  or no  $\theta$  at all. Often, as we will see in Example 3.1, we only care about some function of  $\theta$ . We call this the target parameter,  $\pi(\theta)$ . In this case we may be interested in,

$$\Pi^*(G) \equiv \{\pi(\theta) : \theta \in \Theta^*(G)\}$$

This can sometimes be weaker than finding  $\Theta^*(G)$ . Obviously, if we find the identified set we can determine  $\Pi^*(G)$ . But sometimes we can find  $\Pi^*(G)$  without knowing all of  $\Theta^*(G)$ . In this case our mapping is defined as,  $\phi : \mathcal{G} \rightarrow \mathbb{R}^{d_{\pi(\theta)}}$ .

---

\*This note is partially based on a course taught by Alexander Torgovitsky at the University of Chicago.  
[ariel.boyarsky@yale.edu](mailto:ariel.boyarsky@yale.edu)

## 3. EXAMPLES

It is useful to make some of these idea more concrete with some examples of identification in models.

**EXAMPLE 3.1 (Linear Regression)** *Suppose we observe,  $(Y, X)_{i=1, \dots, n}$  with  $Y \in \mathbb{R}$  and  $X \in \mathbb{R}^p$ . Furthermore suppose we are interested in  $\beta$  such that,*

$$Y = X'\beta + U$$

*Then  $\theta = (\beta, F) \in \Theta$  where  $F$  is the joint distribution of  $(X, U)$ . Then,  $\theta$  clearly implies a distribution for  $(Y, X)$ ,*

$$G_\theta(y, x) = \Pr_F[X'\beta + U \leq y, X \leq x]$$

*Where  $\Pr_F$  is the probability measure under the joint distribution  $F$ . Then we can make the typical assumptions for OLS to restrict the parameter space,*

$$\Theta \equiv \{\theta \equiv (\beta, F) : \mathbb{E}_F[XU] \neq 0, \mathbb{E}_F[X'X]^{-1} \text{ exists}\}$$

*Furthermore, notice that we are typically satisfied with just finding  $\beta$ . Thus we can definite the target parameter as,*

$$\pi(\theta) = \beta$$

*Now that we have denoted the parameter space we can now characterize the identified set. In particular we make the usual argument for OLS. For any  $(\beta, F) \in \Theta$  we must have that,*

$$\beta = \mathbb{E}_F[X'X]^{-1} \mathbb{E}_F[X(X'\beta + U)]$$

*Now suppose,  $\beta \in \Pi^*(G)$ . This implies that there exists a joint distribution,  $F$ , such that  $(\beta, F) \in \Theta^*(G)$ . Then  $\theta$  must satisfy the above equation for  $\beta$  and  $G_{(\beta, F)} = G$  since solving the above equation must mean that we are observationally equivalent to  $G$ . This simply because the best linear predictor will yield residuals,  $U$ , and coefficients,  $\beta$ , such that  $Y = X'\beta + U$ . So it will be impossible to tell apart  $G_\theta$  from  $G$ . But then this directly implies that we have constructed a mapping  $\theta$  such that we have,*

$$\beta = \mathbb{E}_F[X'X]^{-1} \mathbb{E}_F[X(X'\beta + U)] = \mathbb{E}_G[X'X] \mathbb{E}_G[X'Y] \equiv \phi(G)$$

*Thus, we have identified a mapping of  $G$  to the parameter space. Furthermore since  $\phi(\cdot)$  is well defined we will be point identified.*