Recitation Notes: Identification

Prepared by Fei Tan

This recitation note reviews one of the most important notions in econometrics, namely the notion of identification, following the seminal paper Hurwicz (1962). In what follows, we begin with a general definition of identification.

General Definition

[1] In the context of econometrics, identifiability of the unknown quantity (e.g. parameters or functions of parameters) of our interest serves as the *necessary* condition for the existence of a consistent estimator for that quantity.[2] That is, if one could not logically deduce the value of the unknown quantity from a presample analysis, then the usual argument for proving the consistency of an estimator would fail, let alone the derivation of its asymptotic distribution.

To fix idea, we denote by P_X the true distribution of the observed data, which is the probability measure on the state space of random element X. Also, let

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\} \tag{1}$$

be our model for the distribution of the same data, which is indexed by the parameter θ. The model $\mathcal{P}$ is of particular interest because we are intended to interpret it as a *structural* model for the distribution of the observed data that helps us not only summarize various kinds of statistics, but also understand the underlying mechanism that generates the data.[3]

We assume that the model $\mathcal{P}$ is *complete* in the following sense: there exists some $\theta \in \Theta$ such that $P_\theta = P_X$. Cautions are needed here because it is possible that there exists another $\theta^* \in \Theta$ such that $\theta^* \neq \theta$ and $P_{\theta^*} = P_X$ if, for example, the observed data is not sufficiently informative. Therefore, the only legitimate claim we can make from the knowledge of P_X along is that

$$\theta \in \Theta_0(P_X) := \{\theta \in \Theta : P_\theta = P_X\} \tag{2}$$

where $\Theta_0(P_X)$ is referred to as the identified set. We say that θ is *identified* if $\Theta_0(P_X)$ is a singleton.[4] Thus, the natural question is that under what restrictions on $\Theta_0(P_X)$ can the identification of θ be achieved. We will explore these restrictions in the next section.

[1]For more detailed discussion, please see Azeem Shaikh's series of lecture notes on Topics in Econometrics.

[2]The condition is only necessary because the estimator may not converge in a probabilistic sense to the true value of the quantity due, for example, to the existence of excessive dependency in the data. That is, the law of large numbers could fail even when the unknown quantity can be identified.

[3]We will introduce the notion of structural model in the next section. Also, there will be no need to introduce such a model if our interest lies only in the statistical characterization of the observed data. P_X alone would suffice for that purpose.

[4]Identification in our general discussion so far is really a global concept. Rothenberg (1971) showed that the necessary and sufficient condition for *local* identification is the nonsingularity of the Fisher information matrix. Iskrev (2008) shows how the information matrix of linearized DSGE models can be evaluated analytically, which can be used to check the local identifiability in DSGE models.

As an example, consider the following linear regression model

$$Y = X'\beta + \epsilon \tag{3}$$

where X is a $k \times 1$ vector and $\theta = (P_X, \beta, P_{\epsilon|X})$. Then under the following restrictions, θ can be identified:

- A1. $\mathbb{E}_{P_\theta}[\epsilon|X] = 0$.
- A2. There exists no $A \subseteq \mathbb{R}^k$ such that $P_X(A) = 1$ and A is a proper linear subspace of $\mathbb{R}^k$.

The above restrictions may look unfamiliar. A more conventional set of restrictions on $\Theta_0(P)$ under which θ can be identified are the following:

- B1. $\mathbb{E}_{P_\theta}[\epsilon X] = 0$.
- B2. $\mathbb{E}_{P_\theta}[XX']$ is nonsingular.

Think about the proofs. Note that restriction (B1) is usually referred to as the orthogonality condition in the traditional framework of classical regression models. Park (2010) proposes a more general class of regression models, called the *martingale regression* since it is identified by the condition that the error process is a martingale. It allows for the presence of arbitrary time-varying and stochastic volatilities that are often quite persistent and strongly endogenous.[5]

A Priori Information and Prediction

[6] Our general definition of identification has hid two essential points of identification: both the degree of and the need for identification are *relative* notions. First, we describe the identification issue one might encounter even in the simplest case. To fix idea, let $x = [x_1, x_2, \ldots, x_n]'$ denote the state of a configuration that takes the form of simultaneous equations system given by

$$f_i(x) = 0, \qquad i = 1, 2, \ldots, m \tag{4}$$

Hurwicz (1962) refers to f_i as the *behavior pattern* of the i-th component in the configuration. In the context of probabilistic models, the zero's on the right hand side of (4) should be replaced by stochastic disturbance terms so that f_i's take the form of, for example, first-order conditions derived from agents' optimizing behavior in a dynamic stochastic general equilibrium (DSGE) model, or the classical linear regression model given in the previous section.

In what follows, we will focus only on the linear deterministic behavior pattern which, in matrix notation, can be compactly written as

$$Ax - b = 0 \tag{5}$$

[5]See Park (2010), "Martingale Regressions for Conditional Mean Models in Continuous Time".

[6]See Hurwicz (1962), "On the Structural Form of Interdependent Systems".

where the behavior pattern is now completely determined by the $m \times n$ matrix A and the $m \times 1$ vector b. Note that (5) imposes a restriction on the possible values that x can take under the particular configuration specified by (A, b). Let $\mathcal{H}$ be the state space of x that is spanned by the true values of A and b. Following the language and notation established in the previous section, we can take P_X to be $\mathcal{H}$ here because the knowledge of P_X boils down to that of $\mathcal{H}$ in a deterministic setting. Therefore, the identified set of (A, b) can be written as

$$\Theta_0(\mathcal{H}) = \{(A, b) : Ax - b = 0, \quad \forall\ x \in \mathcal{H}\} \tag{6}$$

and the completeness assumption of our model $\mathcal{P}$ amounts to requiring that $\Theta_0(\mathcal{H})$ is a nonempty set. Cautions are needed again because it is straightforward to see that for any $(A, b) \in \Theta_0(\mathcal{H})$ and any $m \times m$ invertible matrix P, the combination

$$C = PA \quad \text{and} \quad d = Pb \tag{7}$$

is also contained in $\Theta_0(\mathcal{H})$. Since these two behavior patterns both span the state space $\mathcal{H}$, we say that (A, b) and (C, d) are *observationally equivalent* with data. Therefore, the knowledge of $\mathcal{H}$ alone fails to single out the true behavior pattern specified by a particular combination of A and b.

The above discussion still remains unclear about how much identification one is able to achieve and what purposes is identification intended for. Indeed, both the degree of and the need for identification are only defined as relative notions that are intimately connected.

1. <u>Identification is no free lunch.</u> Because the knowledge of $\mathcal{H}$ alone does not suffice here, we may look for additional restrictions that are not contained in $\mathcal{H}$ for the purpose of identification. These extra restrictions are usually called *a priori* information, which comes from our economic theory or empirical experience, etc., and can be conveniently formulated in the Bayesian framework by imposing appropriate prior probability measures on (A, b). Thus, the imposition of *a priori* information helps us rule out all those transformation matrices P in (7) that are incompatible with the restrictions provided by the *a priori* information. To see this more clearly, consider the following two extreme cases:

    - If the *a priori* information is sufficiently restrictive about (7) so that the class of all transformation matrices boils down to one containing only diagonal matrices, then we say that the behavioral pattern (A, b) can be identified (up to normalization).
    - If our economic theory or empirical experience remains silent, that is, there is no readily available *a priori* information, then all invertible transformation matrices are allowed and we say that the behavioral pattern (A, b) cannot be identified.

    Therefore, the cardinality of $\mathcal{P}_I$, which is the set of all transformation matrices that are compatible with a given *a priori* information I, measures exactly the degree of identification one is able to achieve *relative* to her *a priori* information I. That is, identification might be quite expensive and I is the price that one must pay for it. But do we always need to pay a high price in order to buy identification?

2. There is price discount on identification. As we shall see, the price of identification really depends on the specific purposes our behavioral patter is intended for. Here we closely follow Hurwicz (1962) and interpret the "need" for identification as the "need for purposes of prediction". This requires a clear distinction between the true old behavioral pattern and all the possible modified behavioral patterns that we are intended to predict. For notational ease, let w be a modifying variable that takes values in its domain $\mathcal{W}$. (Imagine that w represents a particular outcome in our underlying "probability space" $\mathcal{W}$.) Let w^* be the realized historical value of w. Then the true old behavioral pattern that generated $\mathcal{H}$ can be written as the combination of $A^* = A(w^*)$ and $b^* = b(w^*)$. Moreover, all the possible modified behavioral patterns, indexed by elements of W, can be written as

$$A_0(w) = \phi[A^*, b^*, w] \quad \text{and} \quad b_0(w) = \psi[A^*, b^*, w] \tag{8}$$

where both ϕ and ψ are known functions to us. (8) simply says that knowledge of the true old behavioral pattern is sufficient to determine the predicted behavioral pattern indexed by w, although the true old one itself remains undetermined. Now our need for identification of the true old behavioral pattern is for the purpose of finding the set of all predicted states of the configuration over the entire possibilities of our anticipated modifications

$$X_0(\mathcal{W}) = \{x : A_0(w)x - b_0(w) = 0, \quad \forall\ w \in \mathcal{W}\} \tag{9}$$

Similarly, the above set based on alternative old behavioral pattern (not necessarily the true one) can be written as

$$X_P(\mathcal{W}) = \{x : A_P(w)x - b_P(w) = 0, \quad \forall\ w \in \mathcal{W}\} \tag{10}$$

where

$$A_P(w) = \phi[PA^*, Pb^*, w] \quad \text{and} \quad b_P(w) = \psi[PA^*, Pb^*, w]$$

for some transformation matrix P. Clearly, (9) plays a similar role as the *a priori* information I in ruling out all those transformation matrices P in (7) that are incompatible with the restrictions provided by (9), i.e. all P's such that $X_P(\mathcal{W}) \neq X_0(\mathcal{W})$. In much of the structural vector autoregression (SVAR) literature, for example, (9) corresponds to the restrictions imposed on the contemporaneous matrix with the interpretation that there are delays in agents' reactions to the disturbances originating outside of their own sector.[7] To see this more clearly, consider again the following two extreme cases:

- If $\mathcal{W}$ is sufficiently restrictive about (7), then the class of all transformation matrices compatible with (9) boils down to one containing only diagonal matrices.
- If $\mathcal{W} = \{w^*\}$ so that the future configuration remains the same as the old one, then all invertible $m \times m$ transformation matrices are compatible with (9).

[7] See Chris Sims' series of lecture notes on time series econometrics.

Therefore, the elements of $\mathcal{P}_{\mathcal{W}}$, which is the set of all transformation matrices that are compatible with (9), prescribe exactly those behavioral patterns that yield the same prediction. Now it becomes clear that, at least for the purpose of prediction, it suffices to impose *a priori* information enough to eliminate all those transformation matrices outside of $\mathcal{P}_{\mathcal{W}}$ so that

$$\mathcal{P}_{\mathcal{I}} \subseteq \mathcal{P}_{\mathcal{W}} \tag{11}$$

This is the main result of Hurwicz (1962): *the need for knowledge of the true old behavioral pattern is not absolute, but relative to the modification domain.* Therefore, we will always get a price discount for buying identification if our anticipated modifications are not rich enough.

As a final remark of this notes, if $\mathcal{P}_{\mathcal{W}}$ contains only diagonal matrices, then we call (A^*, b^*) the *structure* of the configuration with respect to the modification domain W.