1 Identification in Econometrics

A minimal requirement on an estimator is consistency, i.e., as the sample size increases, the estimator converges in a probabilistic sense to the unknown value of the parameter. We will now study a necessary condition for the existence of consistent estimators. The analysis of identification asks the following question: Can one logically deduce the unknown value of the parameter from the distribution of the observed data? If the answer to this question is "no" under a certain set of assumptions, then consistent estimators cannot exist under the same set of assumptions. If, on the other hand, the answer to this question is "yes" under a certain set of assumptions, then consistent estimators may exist (though further assumptions may be required to get laws of large numbers, central limit theorems, etc. to work).

2 A General Definition of Identification

Let P denote the true distribution of the observed data X. Denote by $\mathbf{P} = \{P_\theta : \theta \in \Theta\}$ a model for the distribution of the observed data. We will assume that $P \in \mathbf{P} = \{P_\theta : \theta \in \Theta\}$. In other words, we assume that the model is correctly specified in the sense that there is some $\theta \in \Theta$ such that $P_\theta = P$. We are interested in θ or perhaps some function f of θ.

Suppose it is known that the distribution of the observed data is $P \in \mathbf{P}$. Since the model is correctly specified by assumption, it is known *a priori* that there exists some $\theta \in \Theta$ such that $P_\theta = P$. But we cannot distinguish any such $\theta \in \Theta$ from any other $\theta^* \in \Theta$ such that $P_{\theta^*} = P$. Thus, from knowledge of P alone, all we can say is that $\theta \in \Theta_0(P)$, where

$$\Theta_0(P) = \{\theta \in \Theta : P_\theta = P\} \ .$$

We will refer to $\Theta_0(P)$ as the identified set. We say that θ is identified if $\Theta_0(P)$ is a singleton for all $P \in \mathbf{P}$.

As mentioned earlier, often we are not interested in θ itself, but rather only a function f of θ. For example, if $\theta \in \mathbf{R}^k$ where $k \geq 2$, then it might be the case that $f(\theta) = \theta_1$. As before, from knowledge of P alone, all we can say is that $f(\theta)$ lies in

$$f(\Theta_0(P)) = \{f(\theta) : \theta \in \Theta_0(P)\} \ .$$

We say that $f(\theta)$ is identified if this set is a singleton for all $P \in \mathbf{P}$.

The model $\mathbf{P}$ here is intended to be interpreted as a "structural" model for the distribution of the observed data. Why should we bother with a structural model? After all, from the distribution of the observed data, P, one can compute all sorts of interesting statistics (e.g., best predictors, best linear predictors, conditional distributions, etc.). The reason is that all of these statistics only describe the data, but do not help us understand the mechanism (the, structure, if you will) that helps generate the data. Here, that structure is embodied by the unknown value of $\theta \in \Theta$ in our model for the data $\mathbf{P}$. The question we seek to answer here is under what conditions is it possible to learn about θ (or some feature of θ) from the distribution of the observed data P.

3 Example 1: Linear Regression

Consider the following linear regression model:

$$Y = X'\beta + \epsilon \ . \tag{1}$$

In this case, $\theta = (P_X, \beta, P_{\epsilon|X})$ and Θ is the set of all possible values for θ. Notice that θ together with the structure of the model determines a unique distribution of the observed data. The following theorem shows that under certain restrictions on Θ, θ is in fact identified.

Theorem 3.1 Suppose for all $\theta \in \Theta$

A1. $E_{P_\theta}[\epsilon|X] = 0$;

A2. There exists no $A \subseteq \mathbf{R}^k$ such that A has probability 1 under P_X and A is a proper linear subspace of $\mathbf{R}^k$.

Then, θ is identified.

PROOF: We need to show that $\Theta_0(P)$ is always a singleton. Let P be given and suppose by way of contradiction that there exists $\theta = (P_X, \beta, P_{\epsilon|X})$ and $\theta^* = (P^*_X, \beta^*, P^*_{\epsilon|X})$ such that $\theta \neq \theta^*$ and $P_\theta = P_{\theta^*} = P$.

First note that since we may recover the marginal distribution of X from the joint distribution of (Y, X), it must be the case that $P_X = P^*_X$.

Second, note that A1 implies that

$$E_{P_\theta}[\epsilon|X] = E_{P_{\theta^*}}[\epsilon|X] = 0 \ .$$

Hence, $E_{P_\theta}[Y|X] = X'\beta$ and $E_{P_{\theta^*}}[Y|X] = X'\beta^*$. Since by assumption $P_\theta = P_{\theta^*}$, it must be the case that $P_X\{X'\beta = X'\beta^*\} = 1$. Assumption A2 implies that this is only possible if $\beta = \beta^*$. To see this, recall the fact that the set $A = \{x \in \mathbf{R}^k : x'(\beta - \beta^*) = 0\}$ is a proper linear subspace of $\mathbf{R}^k$ if $\beta \neq \beta^*$.

Finally, it now follows from (1) that $P_{\epsilon|X} = P^*_{\epsilon|X}$. Thus, $\theta = \theta^*$. ■

We can establish this same conclusion under a slightly different (and more conventional) set of assumptions:

A1′. $E_{P_\theta}[\epsilon X] = 0$;

A2′. $E_{P_\theta}[XX']$ is nonsingular.

4 Example 2: Binary Response Model

The following "threshold-crossing" model of binary response has been applied extensively in economics, medicine and other fields:

$$Y = \mathbf{1}\{X'\beta - \epsilon \geq 0\} \ .$$

In economics, Y usually indicates a utility-maximizing decision maker's observable choice between two alternatives. Then, the latent index $X'\beta - \epsilon$ can be interpreted as the difference in the utility between these two choices.

In medicine, Y is typically an observable binary indicator of health status (e.g., whether or not you are alive). Here, the latent index $X'\beta - \epsilon$ is a measure of health status.

In this case, as before, $\theta = (P_X, \beta, P_{\epsilon|X})$ and Θ is the set of all possible values for θ. To try to make our lives as easy as possible, let's go ahead and make the following restrictions on Θ:

B1. $P_{\epsilon|X} = N(0, \sigma^2)$.

B2. There exists no $A \subseteq \mathbf{R}^k$ such that A has probability 1 under P_X and A is a proper linear subspace of $\mathbf{R}^k$.

Given assumption B1, we may simply write σ in place of $P_{\epsilon|X}$. Let's now see what happens when we try to carry out the same argument used to prove Theorem 3.1. Let P be given and suppose that there exists $\theta = (P_X, \beta, \sigma)$ and $\theta^* = (P_X^*, \beta^*, \sigma^*)$ such that $\theta \neq \theta^*$ and $P_\theta = P_{\theta^*} = P$.

As before, we have immediately that $P_X = P_X^*$.

From B1, we have that $P_\theta\{Y = 1|X\} = \Phi(X'\beta/\sigma)$ and $P_{\theta^*}\{Y = 1|X\} = \Phi(X'\beta^*/\sigma^*)$. Since $P_\theta = P_{\theta^*}$ by assumption, it follows from B2 that

$$\beta/\sigma = \beta^*/\sigma^* \ . \tag{2}$$

We cannot conclude, however, that $\beta = \beta^*$ and $\sigma = \sigma^*$. Indeed, our analysis shows that any θ and θ^* for which (2) holds and $P_X = P_X^*$ satisfies $P_\theta = P_{\theta^*}$. Put differently, even though we cannot identify θ, we can identify $f(\theta) = (P_X, \beta/\sigma)$.

We can summarize the above discussion with the following result:

Theorem 4.1 Suppose all $\theta \in \Theta$ satisfy B1 and B2. Then, $f(\theta) = (P_X, \beta/\sigma)$ is identified.

In practice, people typically assume further that $||\beta|| = 1$, $\beta_1 = 1$ or $\sigma = 1$. Such an assumption, together with assumptions B1 and B2, are enough to identify θ.

4.1 Mean Independence

The above analysis required the rather restrictive parametric assumption B1. In the case of the linear model, it suffices to only assume A1, which is implied by B1, so it is natural to ask whether we can relax B1 in a similar fashion in the binary response model. Concretely, we will replace B1 with the following assumption:

B1′. $E_{P_\theta}[\epsilon|X] = 0$ and $P_{\epsilon|X}$ has support equal to $\mathbf{R}$ with probability 1 under P_X.

Notice that B1 implies B1′.

Unfortunately, the answer to this question is "no", even if we only require the more modest goal of identifying $f(\theta) = \beta$. To see this, consider $\theta = (P_X, \beta, P_{\epsilon|X})$ and any $\beta^* \neq \beta$. We will construct $\theta^* = (P^*_X, \beta^*, P^*_{\epsilon|X})$ satisfying the restrictions on our model and such that $P_{\theta^*} = P_\theta$.

First of all, for $P_{\theta^*} = P_\theta$ to hold, it must be the case that $P_X = P^*_X$, so we are only free to choose $P^*_{\epsilon|X}$. But keep in mind that we must choose it in a way so that B1′ is satisfied. Note that $P_\theta\{Y = 1|X\} = P_{\epsilon|X}\{X'\beta \geq \epsilon\}$. Assumption B1′ implies that this probability lies strictly between (0,1) with probability 1 under P_X. In order to satisfy $P_{\theta^*} = P_\theta$, we must choose $P^*_{\epsilon|X}$ so that it puts mass $P_\theta\{Y = 1|X\}$ less than $X'\beta$. In order to satisfy B1′, we must place the rest of the mass sufficiently high so that it has mean zero. We can do this independently for each value of X. With a bit more thought, you should be able to convince yourself that you can do this in a way that support is equal to $\mathbf{R}$ for each value of X.

4.2 Median Independence

So, it turns out that mean independence is not enough to identify β (even just up to scale) in the binary response model. But mean independence is only one measure of central tendency of a random variable. If we measure central tendency by the median instead of the mean, then it turns out that we can find reasonable conditions under which it is possible to identify β. Specifically, we can get away with the following assumptions on Θ:

C1. $||\beta|| = 1$.

C2. $\text{Med}(\epsilon|X) = 0$ with probability 1 under P_X.

C3. There exists no $A \subseteq \mathbf{R}^k$ such that A has probability 1 under P_X and A is a proper linear subspace of $\mathbf{R}^k$.

C4. P_X is such that at least one component of X has support equal to $\mathbf{R}$ conditional on the other components with probability 1 under P_X. Moreover, the corresponding component of β is nonzero.

Let's compare these assumptions with those used to establish Theorem 4.1. Assumption C1 is needed for exactly the same reason that only β/σ was identified earlier. Assumption C2 is strictly weaker than B1. Assumption C3 is identical to B2. Assumption C4 strengthens the assumption on P_X and also places a mild restriction on β.

The following lemma will help us prove the above result:

Lemma 4.1 Let $\theta = (P_X, \beta, P_{\epsilon|X})$ satisfying C2 be given. Consider any β^*. If $P_\theta\{X'\beta^* < 0 \leq X'\beta \cup X'\beta < 0 \leq X'\beta^*\} > 0$, then there exists no $\theta^* = (P_X^*, \beta^*, P_{\epsilon|X}^*)$ satisfying C2 and also having $P_\theta = P_{\theta^*}$.

Proof: Suppose by way of contradiction that $P_\theta\{X'\beta^* < 0 \leq X'\beta \cup X'\beta < 0 \leq X'\beta^*\} > 0$ yet there exists such a θ^*. As usual, because $P_\theta = P_{\theta^*}$, we have immediately that $P_X = P_X^*$.

Note that $P_\theta\{Y = 1|X\} \geq .5$ if and only if $P_\theta\{X'\beta \geq \epsilon|X\} \geq .5$. By C2, this latter statement is true if and only if $X'\beta \geq 0$. Thus, $P_\theta\{Y = 1|X\} \geq .5$ if and only if $X'\beta \geq 0$.

Likewise, if θ^* satisfies C2, it must be the case that $P_{\theta^*}\{Y = 1|X\} \geq .5$ if and only if $X'\beta^* \geq 0$.

Yet, with positive probability, we have that either $X'\beta^* < 0 \leq X'\beta$ or $X'\beta < 0 \leq X'\beta^*$, which implies that either $P_\theta\{X'\beta \geq \epsilon|X\} < .5 \leq P_{\theta^*}\{Y = 1|X\} \geq .5$ or $P_{\theta^*}\{Y = 1|X\} < .5 \leq P_\theta\{Y = 1|X\}$. This contradicts the fact that $P_\theta = P_{\theta^*}$. ■

With this lemma, we now can prove the following result:

Theorem 4.2 Suppose all $\theta \in \Theta$ satisfy C1, C2, C3 and C4. Then, $f(\theta) = \beta$ is identified.

PROOF: Let $\theta = (P_X, \beta, P_{\epsilon|X})$ satisfying C1, C2, C3 and C4 be given. Assume w.l.o.g. that the component of X specified in C4 is the kth component. Suppose further that $\beta_k > 0$. The same argument *mutatis mutandis* will establish the result for $\beta_k < 0$.

Consider any $\beta^* \neq \beta$. We wish to show that there is no $\theta^* = (P_X^*, \beta^*, P_{\epsilon|X}^*)$ satisfying C1, C2, C3 and C4 and also having $P_\theta = P_{\theta^*}$. From Lemma 4.1, it suffices to show that $P_\theta\{X'\beta^* < 0 \leq X'\beta \cup X'\beta < 0 \leq X'\beta^*\} > 0$. We must consider three cases separately:

(i) Suppose $\beta_k^* < 0$. Then,

$$P_\theta\{X'\beta^* < 0 \leq X'\beta\} = P_\theta\{-\frac{X'_{-k}\beta^*_{-k}}{\beta_k^*} < X_k, -\frac{X'_{-k}\beta_{-k}}{\beta_k} \leq X_k\}.$$

This probability is positive by C4.

(ii) Suppose $\beta_k^* = 0$. Then,

$$P_\theta\{X'\beta^* < 0 \leq X'\beta\} = P_\theta\{X'_{-k}\beta^*_{-k} < 0, -\frac{X'_{-k}\beta_{-k}}{\beta_k} \leq X_k\},$$

$$P_\theta\{X'\beta < 0 \leq X'\beta^*\} = P_\theta\{0 \leq X'_{-k}\beta^*_{-k}, X_k < -\frac{X'_{-k}\beta_{-k}}{\beta_k}\}.$$

Either $P_\theta\{X'_{-k}\beta^*_{-k} < 0\} > 0$ or $P_\theta\{0 \leq X'_{-k}\beta^*_{-k}\} > 0$. If it is the former, then C4 shows that the first of the two above probabilities is positive; if it is the latter, then C4 shows that the second of the two above probabilities is positive.

(iii) Suppose $\beta_k^* > 0$. Then,

$$P_\theta\{X'\beta^* < 0 \leq X'\beta\} = P_\theta\{-\frac{X'_{-k}\beta_{-k}}{\beta_k} \leq X_k < -\frac{X'_{-k}\beta^*_{-k}}{\beta_k^*}\},$$

$$P_\theta\{X'\beta < 0 \leq X'\beta^*\} = P_\theta\{-\frac{X'_{-k}\beta^*_{-k}}{\beta_k^*} < X_k \leq -\frac{X'_{-k}\beta_{-k}}{\beta_k}\}.$$

Assumption C1 implies that β^* is not a scalar multiple of β. Therefore, $\beta^*_{-k}/\beta^*_k \neq \beta_{-k}/\beta_k$. It follows from C3 that $P_\theta\{X'_{-k}\beta^*_{-k}/\beta^*_k \neq X'_{-k}\beta_{-k}/\beta_k\} > 0$. Thus, at least one of the two intervals in the probabilities above must have positive length with positive probability. Assumption C4 thus implies that at least one of these two probabilities must be positive.

This concludes the proof. ■