Asymptotic Theory

Edicson Luna

University of Pennsylvania

Identification

Consistency

Asymptotic distribution

Identification

The 3 steps of good work

1. **Step 1:** Define the effect of interest.
   It implies asking "what if?". It is contrary to just describing ("what is").

2. **Step 2:** Identification of the target parameter.
   Identification links the thought experiment and data.

3. **Step 3:** Statistical inference.
   In practice, we only see a finite sample of the observables. Here, we want to use asymptotic theory to talk about the real parameters.

Essential concept

In general,

- We have a question (target parameter)
- We have a model (assumptions)
- We have some data (we are empiricists)
- We want to use the model and data to answer the question

Identification is about making a *logically coherent* empirical conclusion.

Identification

In this class, our main objective will be identifying the parameter β in the linear regression $Y = X\beta$ (considering one or more explanatory variables).

But what is identification? In simple words, it is to be able to express the parameter in terms of the data after we assume a model (in this case, a linear model). But let's add a little formality and much more intuition.

Formal definition

Let P denote the true distribution of the observed data $(y_i, x_i)_{i=1}^n$. Denote by $\mathbf{P} = \{P_\theta : \theta \in \Theta\}$ a model for the distribution of the observed data. We assume $P \in \mathbf{P}$. In other words, we assume $\exists \theta \in \Theta$ such that $P_\theta = P$. We are interested in θ. (Θ is the set of all possible values of θ).

We define the set $\Theta_0(P) := \{\theta \in \Theta : P_\theta = P\}$ which is called "the identified set". We sat that θ is identified if $\Theta_o(P)$ is a singleton $\forall P \in \mathbf{P}$.

More on identification

Notice that identification has nothing to do with statistical inference.

$$\text{sample} \rightarrow \text{population} \rightarrow \text{unobserved parameters}$$

Identification is about the second arrow. Statistical inference is about using the sample to learn about the population (we will be back to this soon).

The second arrow is logically the first thing to consider: can we recover the population parameter when we know the population distribution?

Why asymptotics?

Notice that when we analyzed the expectation of our estimators, we used finite sample properties. Now we will work with large sample properties.

But, why are we working with asymptotic properties if we (almost) always have finite data? The reason is that obtaining unbiased estimators is usually not possible. That is why economists typically focus on what would happen if $n \to \infty$.

Economists agree that consistency (I will define it soon) is the minimal requirement for an estimator.

Consistency

Consistency: formal definition

Let $\hat{\theta}$ be an estimator of θ based on some data $(X_i)_{i=1}^n$. Then, $\hat{\theta}$ is a **consistent** estimator of θ if $\forall \epsilon > 0$,

$$\mathbb{P}(|\hat{\theta} - \theta| > \epsilon) \to 0 \quad \text{as } n \to \infty$$

We also say that $\hat{\theta}$ **converges in probability** to θ. If $\hat{\theta}$ is not consistent, we say that the estimator is **inconsistent**.

Notice that if we have a consistent estimator, $\hat{\theta} \xrightarrow{p} \theta$. This means that **having consistency implies that we can identify the parameter** (which is our final goal). That is why we want to talk about consistency.

Theorems

Remember that previously we derived $\beta = [\mathrm{Var}(X)]^{-1}\mathrm{Cov}(X, Y)$. The LLN and CLT we studied before are not completely useful for what comes now. The reason is the following.

- Consider a sequence $X_1, X_2, ..., X_n$ of RVs that are i.i.d. Now, fix n and define a sequence of RVs defined as $W_2, ..., W_n$ such that $W_i = (X_i - \bar{X}_n)^2$ where $\bar{X}_n$ is the average of the first n X's.

Theorems

- For simplicity compare W_2 and W_3. Notice that in both cases you have the term $\bar{X}_n$.
- Therefore, there is not complete independence between W_2 and W_3.
- And so, since there is a degree of dependence in the sequence $W_1, W_2, ...$ (they are not i.i.d.), we can not use our theorems. The good news is that there are some relaxed versions of the LLN and the CLT.

Relaxed theorems

Assume there is weak dependence in a sequence of RVs $W_1, W_2, \ldots$

- **LLN (V2):** As long as $E[W_i^2] < \infty$, it is true that $\bar{W}_n \xrightarrow{p} E[W_i]$ as $n \to \infty$.
- **CLT (V2):** As long as $E[W_i^{2+\delta}] < \infty$ for $\delta > 0$, it is true that $\frac{\bar{W}_n - E[W_i]}{\text{se}(\bar{W}_n)} \xrightarrow{d} Z$ as $n \to \infty$. (Z is the standard normal distribution). This is the same as saying $\sqrt{n}(\bar{W}_n - E[W_i]) \xrightarrow{d} \mathcal{N}(0, V)$.

CLM Assumptions

Let's quickly remember our CLM assumptions (in the model with more than one regressor)

1. Linearity in parameters. That is, the true relationship between y_i and x_i is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_{k-1} x_{ik-1} + \epsilon_i$$

2. There is random sampling. That is, the observations are i.i.d.
3. The matrix $\mathbb{E}\left[X'X\right]$ has complete rank. (i.e. it's invertible).
4. Zero conditional mean, $\mathbb{E}[\epsilon|X] = 0$. This also may be said as $\mathbb{E}[\epsilon_i|x_i] = 0$

Consistency of the estimator

Theorem. Under assumptions 1 to 4, the OLS estimator is consistent.

Proof. Remember

$$
\begin{aligned}
\hat{\beta} = (X'X)^{-1}X'Y &= \beta + (X'X)^{-1}X'\epsilon \\
&= \beta + \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i \epsilon_i \\
&= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i \epsilon_i
\end{aligned}
$$

Consistency of the estimator

By the LLN, assuming weak dependence

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{p} \mathbb{E}[x_i x_i']$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i e_i \xrightarrow{p} \mathbb{E}[x_i \epsilon_i]$$

Notice that $\mathbb{E}[x_i e_i] = 0$ by the LIE. Then (by Slutsky's theorem)

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{p} \mathbb{E}[x_i x_i']^{-1} \times 0 = 0$$

Relaxation of the assumption

Notice that $\mathbb{E}[\epsilon_i|x_i] = 0$ implies $\mathbb{E}[x_i\epsilon_i] = 0$, but not the contrary. That is, we could perfectly relax $\mathbb{E}[\epsilon_i|x_i] = 0$ and assume that $\mathbb{E}[x_i\epsilon_i] = 0$ (which is a weaker assumption since it is implied by the former).

Slutsky is about being able to multiply the expectations in the limit. We will use this concept again when we work with the asymptotic distribution.

Slutsky for consistency (V1)

There are two versions of Slutsky. So far, let's state the one we just used.

Slutsky's theorem. Let $(X_n)_{n=1}^N$ and $(Y_n)_{n=1}^N$ two sequences of RVs. Take c as a constant. If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} c$, it is true that

- $X_n + Y_n \xrightarrow{p} X + c$
- $X_n Y_n \xrightarrow{p} Xc$
- If $c \neq 0$, $\frac{X_n}{Y_n} \xrightarrow{p} \frac{X}{c}$

Notice that we used the one in the middle.

Asymptotic distribution

Motivation

Notice that consistency is important since it tells us that our estimator goes in the right direction. We at least know that as we add observations (which should be i.i.d.) we are getting closer and closer to the real parameter (i.e. $\hat{\beta}$ is converging into β as n increases). As mentioned before, consistency is the minimum requirement for any estimator.

Nonetheless, notice that consistency does not allow us to perform statistical inference. That is why, now we focus on the sampling distribution of the OLS estimators.

Assumptions

We need to bring back assumption 5.

5. The error is constant given any value of X.

$$\text{Var}(\epsilon|X) = \sigma^2 I_n$$

Notice that previously we stated a 6th assumption about ϵ_i being normally distributed (and therefore y_i). Now, we will not need it. The reason as you may expect is that even if ϵ_i is not normally distributed, we will have normality by the CLT.

Assymptotic normality of OLS

Theorem. Under assumptions 1 to 5,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, V)$$

where $V := \Sigma^{-1}\Omega\Sigma^{-1}$ with $\Sigma = \mathbb{E}[x_i x_i']$ and $\Omega = \mathbb{E}[X'\epsilon\epsilon' X]$

Assymptotic normality of OLS

Proof. Remember $\hat{\beta} = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i$ which is the same as $\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i y_i$. Replacing y_i is not difficult to see

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i \epsilon_i$$

By the CLT

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(X_i e_i) = \mathbb{E}[X' \epsilon \epsilon' X] = \Omega)$$

Assymptotic normality of OLS

Finally, by Slutsky (V2)
$\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{d} \mathcal{N}(0, \Sigma^{-1}\Omega\Sigma^{-1})$.

In conclusion,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, V)$$

Notice that in this case, we did not require the errors to have a normal distribution (as we did in hypothesis testing). The reason, as you already know, is the CLT which gives us the normality if $n \to \infty$. The Slutsky used here is another one which I will define in the next slide.

Convergence in distribution

- **Convergence in distribution:** A sequence $(X_n)_{n=1}^{\infty}$ or RVs, with CDFs $(F_n)_{n=1}^{\infty}$. We say $X_n \xrightarrow{d} X$ if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

Slutsky for asymptotic distribution (V2)

Slutsky's theorem. Let $(X_n)_{n=1}^N$ and $(Y_n)_{n=1}^N$ two sequences of RVs. Take c as a constant. If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$, it is true that

- $X_n + Y_n \xrightarrow{d} X + c$
- $X_n Y_n \xrightarrow{d} Xc$
- If $c \neq 0$, $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$

Again, in the proof we followed the one in the middle.

Natural estimator for variance

To carry out inference, the proposed consistent estimator for the asymptotic variance is of the form

$$\hat{V} := \hat{\Sigma}^{-1}\hat{\Omega}\hat{\Sigma}^{-1}$$

where

$$\hat{\Omega} := X'\epsilon\epsilon' X = \frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i x_i'$$

$$\hat{\Sigma} := \frac{1}{n}\sum_{i=1}^{n} x_i x_i'$$

where e_i is the residual of the regression.

Natural estimator for variance

It is evidently that $\hat{\Sigma} \xrightarrow{p} \Sigma$ by the LLN. Then, as long as $\hat{\Omega} \xrightarrow{p} \Omega$, we use Slutsky (V1) and conclude that $\hat{V} \xrightarrow{p} V$.

$$
\begin{aligned}
\hat{\Omega} &= \frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i x_i' = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i'\hat{\beta}\right)^2 x_i x_i' \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\epsilon_i - x_i'(\hat{\beta}-\beta)\right)^2 x_i x_i' \\
&= \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2 x_i x_i' - 2\frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}-\beta)' x_i \epsilon_i x_i x_i' + \frac{1}{n}\sum_{i=1}^{n}\left[(\hat{\beta}-\beta)' x_i\right]^2 x_i x_i' \\
&\xrightarrow{p} \Omega
\end{aligned}
$$

Given that the second and third terms converge to zero (remember that $\hat{\beta}$ is a consistent estimator for β) and the first term converges to Ω by the LLN.