Standard Errors

Christopher Taber

University of Wisconsin

December 6, 2016

One thing we need to do as empirical economists is to calculate standard errors.

Often a nice way to do this is using the bootstrap which has good properties and is also very easy to program

However, for complicated structural models this is not feasible.

I want to review asymptotic theory for you emphasizing the main ideas and abstracting from some details.

The most Important Asymptotic Theory You Really Need to Know...

Really the key to understand standard errors for most models requires just four things

1. Slutsky
2. Law of Large Numbers
3. Central Limit Theorem
4. Mean Value Theorem

The mean value theorem

As long as f is continuously differentiable for any $a \leq b$ there exists a c

$$a \leq c \leq b$$

such that

$$f'(c) = \frac{f(b) - f(a)}{a - b}$$

The Delta Method

Suppose you know that $\widehat{\theta}$ is approximately normal with known variance covariance matrix

How do you calculate the distribution of $f\left(\widehat{\theta}\right)$?

Its just the mean value theorem with $\widehat{\theta}$ and the true value θ_0

where $\widehat{\theta}$ is a consistent estimate of θ_0 and

$$\sqrt{N}\left(\widehat{\theta}-\theta_0\right) \approx N\left(0, \Sigma_\theta\right)$$

Then we know that for some $\widetilde{\theta}$ between $\widehat{\theta}$ and θ_0 (which means element by element they are in between)

$$f\left(\widehat{\theta}\right) - f(\theta_0) = \frac{\partial f\left(\widetilde{\theta}\right)}{\partial \theta'}\left(\widehat{\theta} - \theta_0\right)$$

throw a $\sqrt{N}$ in front of both sides and you get

$$\begin{aligned}\sqrt{N}\left(f\left(\widehat{\theta}\right) - f(\theta_0)\right) &= \frac{\partial f\left(\widetilde{\theta}\right)}{\partial \theta'}\sqrt{N}\left(\widehat{\theta} - \theta_0\right) \\ &\approx \frac{\partial f\left(\theta_0\right)}{\partial \theta'}\sqrt{N}\left(\widehat{\theta} - \theta_0\right)\end{aligned}$$

but since

$$\sqrt{N}\left(\widehat{\theta}-\theta_0\right) \approx N\left(0,\Sigma_\theta\right)$$

this means that

$$\sqrt{N}\left(f\left(\widehat{\theta}\right)-f(\theta_0)\right) \approx N\left(0,\frac{\partial f\left(\theta_0\right)}{\partial \theta'}\Sigma_\theta\frac{\partial f\left(\theta_0\right)}{\partial \theta}\right)$$

we approximate $\frac{\partial f(\theta_0)}{\partial \theta'}$ with $\frac{\partial f\left(\widehat{\theta}\right)}{\partial \theta'}$

GMM

Lets see how GMM works focusing on the just identified case

The model is

$$E\left[m\left(X_i,\theta_0\right)\right]=0$$

We estimate it by setting

$$0=\frac{1}{N}\sum_{i=1}^{n}m(X_i,\widehat{\theta})$$

To see consistency we use mean value theorem putting $\sqrt{N}$ on both sides

$$\begin{aligned}
&\sqrt{N}\frac{1}{N}\sum_{i=1}^{N} m(X_i,\widehat{\theta}) - \sqrt{N}\frac{1}{N}\sum_{i=1}^{N} m(X_i,\theta_0) \\
&= \left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial m(X_i,\widetilde{\theta})}{\partial\theta'}\right]\sqrt{N}\left[\widehat{\theta}-\theta_0\right]
\end{aligned}$$

$$\sqrt{N}\frac{1}{N}\sum_{i=1}^{N} m(X_i,\widehat{\theta}) = 0$$

by definition of $\widehat{\theta}$

For the second one just think of $m(X_i,\theta_0)$ as an i.i.d. random variable with expected value zero

We can use a central limit theorem on this and we know that

$$\sqrt{N}\frac{1}{N}\sum_{i=1}^{N} m(X_i,\theta_0) \approx N\left(0, E\left[m(X_i,\theta_0)m(X_i,\theta_0)'\right]\right)$$

Now just notice that

$$\begin{aligned}\sqrt{N}\left[\widehat{\theta}-\theta_0\right] =& \left[\frac{1}{N}\sum_{i=1}^{n}\frac{\partial m(X_i,\widetilde{\theta})}{\partial\theta'}\right]^{-1}\sqrt{N}\frac{1}{N}\sum_{i=1}^{N} m(X_i,\theta_0)\\ \approx& N\left(0, G^{-1}\Psi G'^{-1}\right)\end{aligned}$$

where

$$\begin{aligned}G \equiv& E\left(\frac{\partial m(X_i,\theta_0)}{\partial\theta'}\right)\\ \Psi \equiv& E\left[m(X_i,\theta_0)m(X_i,\theta_0)'\right]\end{aligned}$$

In practice you just replace population values with sample analogues

M-Estimation

Think about choosing $\widehat{\theta}$ to maximize

$$\frac{1}{N}\sum_{i=1}^{N} g(X_i,\theta)$$

Examples:

- Maximum Likelihood
- Ordinary Least Squares
- Nonlinear Least Squares

From first order condition we know that

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\partial g(X_i,\widehat{\theta})}{\partial \theta'} = 0$$

But now this is (locally) just like GMM

Just use

$$m(X_i,\theta) = \frac{\partial g(X_i,\theta)}{\partial \theta'}$$

and then we get

$$\sqrt{N}\left[\widehat{\theta}-\theta_0\right] = \left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2 g(X_i,\widetilde{\theta})}{\partial\theta\partial\theta'}\right]^{-1}\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{\partial g(X_i,\theta)}{\partial\theta'}$$

$$\approx N\left(0,\left[E\left(\frac{\partial^2 g(X_i,\theta_0)}{\partial\theta\partial\theta'}\right)\right]^{-1}E\left[\frac{\partial g(X_i,\theta_0)}{\partial\theta}\frac{\partial g(X_i,\theta_0)}{\partial\theta'}\right]\left[E\left(\frac{\partial^2 g(X_i,\theta_0)}{\partial\theta'\partial\theta}\right)\right]^{-1}\right)$$

For maximum likelihood

$$E\left(\frac{\partial^2 g(X_i,\theta_0)}{\partial\theta\partial\theta'}\right) = E\left[\frac{\partial g(X_i,\theta_0)}{\partial\theta}\frac{\partial g(X_i,\theta_0)}{\partial\theta'}\right]$$

Indirect Inference

I want to sketch how this works. it can get more complicated.

From our model

$$
\begin{aligned}
Y_i \equiv & y(X_i, u_i; \theta) \\
u_i \sim & \Psi(u_i; \theta) \\
\widehat{\beta} \equiv & argmin_\beta F\left(\frac{1}{N}\sum_{i=1}^{N} g(X_i, Y_i, \beta), \beta\right) \\
\widetilde{B}(\theta) \equiv & \frac{1}{H}\sum_{h=1}^{H} argmin_\beta F\left(\frac{1}{S}\sum_{s=1}^{S} g(X_{hs}, y(X_{hs}, u_{hs}; \theta); \beta), \beta\right) \\
\widehat{\theta} = & argmin_\theta \left(\widetilde{B}(\theta) - \widehat{\beta}\right)' \Omega \left(\widetilde{B}(\theta) - \widehat{\beta}\right)
\end{aligned}
$$

and some notation:

$$
\begin{aligned}
\beta_0 \equiv & argmin_\beta F(E\left[g(X_i, Y_i, \beta)\right], \beta). \\
B(\theta) = & argmin_\beta F\left(G(\theta, \beta)\right).
\end{aligned}
$$

Without getting into details of how we can show that

$$
\sqrt{N}\left(\widehat{\beta} - \beta_0\right) \sim N\left(0, \Sigma\right)
$$

and notice that if we have the right data generation process

$$
\sqrt{S}\left(\widetilde{B}(\theta) - \beta_0\right) \sim N\left(0, \frac{1}{H}\Sigma\right)
$$

Lets assume these guys are independent of each other (thats not quite right because they both depend on the sample draw of X_i)

The first order condition is

$$0 = -2\frac{\partial \widetilde{B}(\widehat{\theta})'}{\partial \theta}\Omega\left(\widetilde{B}(\widehat{\theta}) - \widehat{\beta}\right)$$

Now we want to use the mean value theorem with

$$f(\theta) = \frac{\partial \widetilde{B}(\theta)'}{\partial \theta}\Omega\left(\widetilde{B}(\theta) - \widehat{\beta}\right)$$

so

$$\sqrt{N}\left(f\left(\widehat{\theta}\right) - f(\theta_0)\right) = \frac{\partial f\left(\widetilde{\theta}\right)}{\partial \theta'}\sqrt{N}\left(\widehat{\theta} - \theta_0\right)$$

First note that

$$
\begin{aligned}
\sqrt{N} f\left(\widehat{\theta}\right) =& 0 \\
\sqrt{N} f(\theta_0) =& \frac{\partial \widetilde{B}(\theta_0)'}{\partial \theta} \Omega \sqrt{N} \left(\widetilde{B}(\theta_0) - \widehat{\beta}\right) \\
=& \frac{\partial \widetilde{B}(\theta_0)'}{\partial \theta} \Omega \left[\sqrt{N} \left(\widetilde{B}(\theta_0) - \beta_0\right) - \sqrt{N} \left(\widehat{\beta} - \beta_0\right)\right] \\
\sim& N\left(0, \frac{\partial \widetilde{B}(\theta_0)'}{\partial \theta} \Omega V \Omega \frac{\partial \widetilde{B}(\theta_0)}{\partial \theta'}\right)
\end{aligned}
$$

To derive V note that

$$\begin{aligned}\sqrt{N}\left(\widetilde{B}(\theta)-\beta_0\right) =& \frac{\sqrt{N}}{\sqrt{S}}\sqrt{S}\left(\widetilde{B}(\theta)-\beta_0\right)\\ \sim& N\left(0, \frac{N}{SH}\Sigma\right)\end{aligned}$$

so

$$V = \left[\frac{N}{SH}+1\right]\Sigma$$

and

$$
\begin{aligned}
\frac{\partial f\left(\widetilde{\theta}\right)}{\partial \theta'} \approx & \frac{\partial \frac{\partial \widetilde{B}(\theta_0)}{\partial \theta'} \Omega \left(\widetilde{B}(\theta_0) - \beta_0\right)}{\partial \theta} \\
= & o_p(1) + \frac{\partial \widetilde{B}(\theta_0)'}{\partial \theta} \Omega \frac{\partial \widetilde{B}(\theta_0)}{\partial \theta'}
\end{aligned}
$$

Thus

$$\sqrt{N}\left(\widehat{\theta}-\theta_0\right) \sim$$
$$N\left(0, \left[\frac{\partial B(\theta_0)'}{\partial\theta}\Omega\frac{\partial B(\theta_0)}{\partial\theta'}\right]^{-1} \frac{\partial\widetilde{B}(\theta_0)'}{\partial\theta}\Omega V\Omega\frac{\partial\widetilde{B}(\theta_0)}{\partial\theta'} \left[\frac{\partial B(\theta_0)'}{\partial\theta}\Omega\frac{\partial B(\theta_0)}{\partial\theta'}\right]^{-1}\right)$$

Two Stage Estimation

Now suppose we estimate a model in two steps

- First estimate θ_1
- Then use estimate of θ_1 to get θ_2

In getting standard errors for θ_2 we need to worry about the fact that θ_1 was estimated

Lets think about how to get our standard errors

Lets focus on the case where both stages are GMM estimators

Stage 1:

Choose $\widehat{\theta}_1$ to solve

$$\frac{1}{N_1}\sum_{i_1=1}^{N_1} m_1(X_{i_1}, \widehat{\theta}_1) = 0$$

Stage 2: Use $\widehat{\theta}_1$ and solve for $\widehat{\theta}_2$ by minimizing

$$\left[\frac{1}{N_2}\sum_{i_2=1}^{N_2} m_2\left(X_{i_2}, \widehat{\theta}_1, \theta_2\right)\right]' W_2 \left[\frac{1}{N_2}\sum_{i_2=1}^{N_2} m_2\left(X_{i_2}, \widehat{\theta}_1, \theta_2\right)\right]$$

This is very common

The standard errors for $\widehat{\theta}_1$ are fine, but the standard errors for $\widehat{\theta}_2$ need to be adjusted to account for the fact that $\widehat{\theta}_1$ is estimated

In terms of notation define for ℓ and j (which can be 1 or 2)

$$G_{\ell j} = E\left(\frac{\partial m_\ell(X_i, \theta_0)}{\partial \theta_j'}\right)$$

and

$$\Psi_{\ell j} = E\left[m_\ell(X_i, \theta_{10}) m_j(X_i, \theta_{10})'\right]$$

Write the first stage using as

$$\sqrt{N}\left[\widehat{\theta}_1 - \theta_{10}\right] = \left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial m_1(X_i, \widetilde{\theta}_1)}{\partial \theta_1'}\right]^{-1}\sqrt{N}\frac{1}{N}\sum_{i=1}^{N} m_1(X_i, \theta_{10})$$

$$\approx G_{11}'^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} m_1(X_i, \theta_{10})$$

For the second stage we use the mean value theorem once again

$$
\begin{aligned}
&\sqrt{N}\frac{1}{N}\sum_{i=1}^{N} m_2(X_i,\widehat{\theta}) - \sqrt{N}\frac{1}{N}\sum_{i=1}^{N} m_2(X_i,\theta_0) \\
&= \left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial m_2(X_i,\widetilde{\theta})}{\partial\theta'}\right]\sqrt{N}\left[\widehat{\theta}-\theta_0\right] \\
&\approx G_{21}\sqrt{N}\left[\widehat{\theta}_1-\theta_{10}\right] + G_{22}\sqrt{N}\left[\widehat{\theta}_2-\theta_{20}\right] \\
&= G_{21}G_{11}'^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} m_1(X_i,\theta_{10}) \\
&\quad + G_{22}\sqrt{N}\left[\widehat{\theta}_2-\theta_{20}\right]
\end{aligned}
$$

And as above

$$\sqrt{N}\frac{1}{N}\sum_{i=1}^{N} m_2(X_i,\widehat{\theta}) =0$$

so

$$\begin{aligned}\sqrt{N}\left[\widehat{\theta}_2 - \theta_{20}\right] =& - G_{22}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[m_2(X_i,\theta_0) + G_{21}G_{11}'^{-1}m_1(X_i,\theta_{10})\right] \\ \sim& N\left(0, G_{22}^{2}\Psi G_{22}\right)\end{aligned}$$

where

$$\begin{aligned}\Psi =& E\left(\left[m_2(X_i,\theta_{10},\theta_{20}) + G_{21}G_{11}'^{-1}m_1(X_i,\theta_{10})\right]\left[m_2(X_i,\theta_{10},\theta_{20}) + G_{21}G_{11}'^{-1}m_1(X_i,\theta_{10})\right]'\right) \\ =& \Psi_{22} + G_{21}G_{11}'^{-1}\Psi_{12} + \Psi_{21}G_{11}^{-1}G_{12} + G_{21}G_{11}'^{-1}\Psi_{11}G_{11}^{-1}G_{12}\end{aligned}$$

A much easier way to do it

Now back to the two step procedure

We can just think of this as one big GMM expression with

$$m\left(X_i, \theta\right) \equiv \left[\begin{array}{c} m_1(X_i, \theta_1) \\ m_2(X_i, \theta_1, \theta_2) \end{array}\right]$$

it is numerically identical and we can use the standard GMM formula to get standard errors

Note that we have written this as GMM, but it also can be M-estimation as we showed above

An alternative way to deal with this is to bootstrap

We have to bootstrap the whole procedure, not just the second stage and we are done