Nonstandard Inference – Set Identification

Frank Schorfheide
University of Pennsylvania
Econ 722 – Part 1

January 22, 2019

Recall: Linear Regression / AR Models

- Consider AR(1) model:

$$y_t = y_{t-1}\phi + u_t, \quad u_t \sim iidN(0,1).$$

- Let $x_t = y_{t-1}$. Write as

$$y_t = x_t'\phi + u_t, \quad u_t \sim iidN(0,1),$$

  or

$$Y = X\phi + U.$$

  We can easily allow for multiple regressors. Assume ϕ is $k \times 1$.

- Notice: we treat the variance of the errors as know. The generalization to unknown variance is straightforward but tedious.

- Likelihood function:

$$p(Y|\phi) = (2\pi)^{-T/2} \exp\left\{ -\frac{1}{2}(Y - X\phi)'(Y - X\phi) \right\}.$$

A Convenient Prior

- Prior:

$$\phi \sim N\left(0_{k\times 1}, \tau^2 \mathcal{I}_{k\times k}\right), \quad p(\phi) = (2\pi\tau^2)^{-k/2} \exp\left\{-\frac{1}{2\tau^2}\phi'\phi\right\}$$

- Large τ means diffuse prior.
- Small τ means tight prior.

Posterior Distribution

- Recall: posterior distribution of ϕ is a multivariate normal distribution

$$\phi|Y \sim N(\bar{\phi}_T, \bar{V}_T)$$

with

$$\begin{aligned} \bar{\phi}_T &= (X'X + \tau^{-2}\mathcal{I})^{-1}X'Y \\ \bar{V}_T &= (X'X + \tau^{-2}\mathcal{I})^{-1}. \end{aligned}$$

- Suppose that

$$\frac{1}{T}\sum_{t=1}^{T} x_t x_t' \stackrel{p}{\longrightarrow} \mathbb{E}[x_t x_t'].$$

- Let $\hat{\phi} = (X'X)^{-1}X'Y$. Then,

$$\sqrt{T}(\phi - \hat{\phi}) \,|\, Y \Longrightarrow N\left(0, \left(\mathbb{E}[x_t x_t']\right)^{-1}\right)$$

Let's Take a Look at the Frequentist Calculations

- Consider model $Y = X\phi + U$.
- OLS estimator: $\hat{\phi} = (X'X)^{-1}X'Y$.
- If model is correct, then

$$\hat{\phi} = \phi + (X'X)^{-1}X'U = \phi + \left(\frac{1}{T} \sum_{t=1}^{T} x_t x_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} x_t u_t \right).$$

- In the "regular" case (recall that $\mathbb{V}[u_1] = 1$):

$$\frac{1}{T} \sum_{t=1}^{T} x_t x_t' \stackrel{p}{\longrightarrow} \mathbb{E}[x_t x_t'], \quad \sqrt{T} \frac{1}{T} \sum_{t=1}^{T} x_t u_t \Longrightarrow N\big(0, \mathbb{E}[x_t x_t']\big).$$

- Thus,

$$\sqrt{T}(\hat{\phi} - \phi) \,|\, \phi \Longrightarrow N\left(0, \big(\mathbb{E}[x_t x_t]\big)^{-1}\right)$$

What Do We Need for the Convergence?

- Suppose that we have a stationary AR(1) model:

$$y_t = \phi_0 y_{t-1} + u_t, \quad |\phi| < 1.$$

- Then,

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{p} \mathbb{E}[y_{t-1}^2] = \frac{1}{1-\phi_0^2}, \quad \sqrt{T}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}u_t \Longrightarrow N\left(0, \frac{1}{1-\phi_0^2}\right).$$

- **Frequentist result:**

$$\sqrt{T}(\hat{\phi} - \phi_0) \,|\, \phi \Longrightarrow N\left(0, 1-\phi_0^2\right).$$

- **Bayesian result:**

$$\sqrt{T}(\phi - \hat{\phi}) \,|\, Y \Longrightarrow N\left(0, 1-\phi_0^2\right).$$

- Properly re-scaled sampling and posterior distributions are asymptotically the same.
- This result holds in many settings that asymptotically look like Gaussian location shift experiments.

Set Identification

- Suppose that y_t is determined by the AR(1) model but object of interest is θ, which can be bounded based on ϕ:

$$\phi \leq \theta \quad \text{and} \quad \theta \leq \phi + 1.$$

- Parameter θ is set-identified.
- The interval $\Theta(\phi) = [\phi, \ \phi + 1]$ is called the identified set.
- Prior for θ conditional on ϕ of the form

$$\theta|\phi \sim U[\phi, \phi + 1].$$

Set Identification – Bayesian Calculations

- Joint posterior of θ and ϕ:

$$p(\theta, \phi|Y) = p(\phi|Y)p(\theta|\phi, Y) \propto p(Y|\phi)p(\theta|\phi)p(\phi).$$

- Since θ does not enter the likelihood function, we deduce that

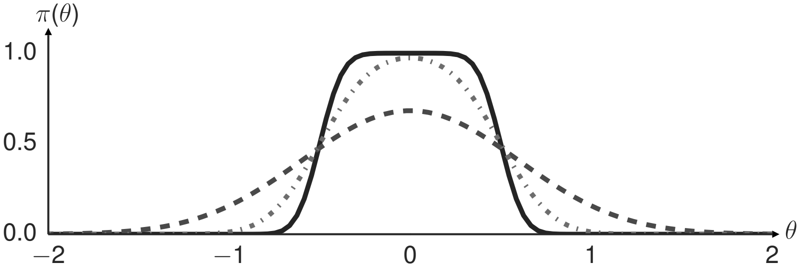
$$\begin{aligned} p(\phi|Y) &= \frac{p(Y|\phi)p(\phi)}{\int p(Y|\phi)p(\phi)d\phi} \\ p(\theta|\phi, Y) &= p(\theta|\phi). \end{aligned}$$

- In our example the marginal posterior distribution of θ is given by

$$\begin{aligned} \pi(\theta) &= \int_{\theta-1}^{\theta} p(\phi|Y)p(\theta|\phi)d\phi \\ &= \Phi_N\left(\frac{\theta-\bar{\phi}}{\sqrt{\bar{V}}}\right) - \Phi_N\left(\frac{\theta-1-\bar{\phi}}{\sqrt{\bar{V}}}\right), \end{aligned}$$

where $\Phi_N(x)$ is the cumulative density function of a $N(0,1)$.

Posterior Density of Set-Identified Model



Posterior distribution $\pi(\theta)$ for $\bar{\phi} = -0.5$ and $\bar{V}_\phi$ equal to $1/4$ (dotted), $1/20$ (dashed), and $1/100$ (solid).

Review: Importance Sampling

1. For $i = 1$ to N, draw $\theta^i \stackrel{iid}{\sim} g(\theta)$ and compute the unnormalized importance weights

$$w^i = w(\theta^i) = \frac{f(\theta^i)}{g(\theta^i)}.$$

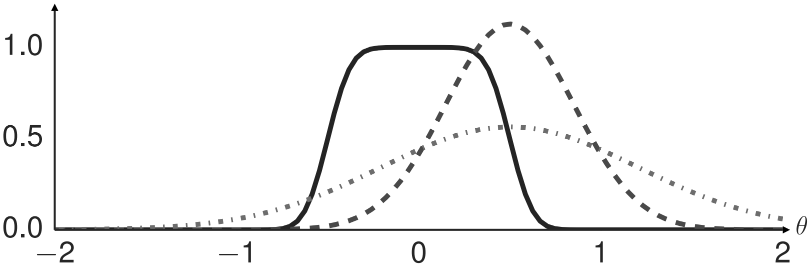
2. Compute the normalized importance weights

$$W^i = \frac{w^i}{\frac{1}{N}\sum_{i=1}^N w^i}.$$

An approximation of $\mathbb{E}_\pi[h(\theta)]$ is given by

$$\bar{h}_N = \frac{1}{N}\sum_{i=1}^N W^i h(\theta^i).$$

Importance Sampling Distribution



Posterior density $\pi(\theta)$ (solid) as well as two importance sampling densities ("concentrated" (dashed) and "diffuse" (dotted)) $g(\theta)$.

Accuracy

- Since we are generating *iid* draws from $g(\theta)$, it's fairly straightforward to derive a CLT:
- It can be shown that

$$\sqrt{N}(\bar{h}_N - \mathbb{E}_\pi[h]) \Longrightarrow N\big(0, \Omega(h)\big), \quad \text{where} \quad \Omega(h) = \mathbb{V}_g[(\pi/g)(h - \mathbb{E}_\pi[h])].$$

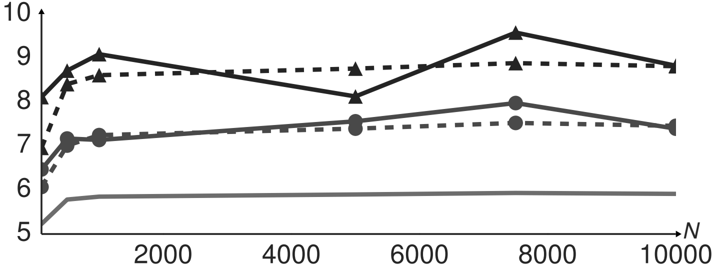
- Using a crude approximation (see, e.g., Liu (2008)), we can factorize $\Omega(h)$ as follows:

$$\Omega(h) \approx \mathbb{V}_\pi[h]\big(\mathbb{V}_g[\pi/g] + 1\big).$$

  The approximation highlights that the larger the variance of the importance weights, the less accurate the Monte Carlo approximation relative to the accuracy that could be achieved with an *iid* sample from the posterior.
- Users often monitor

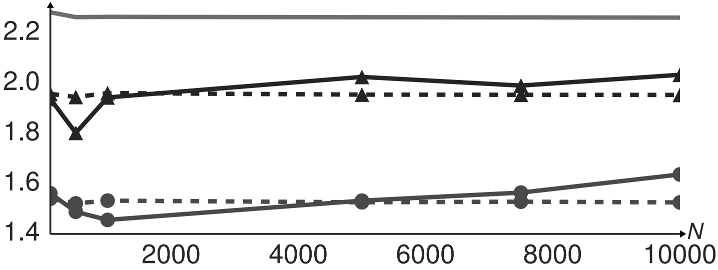
$$ESS = N\frac{\mathbb{V}_\pi[h]}{\Omega(h)} \approx \frac{N}{1 + \mathbb{V}_g[\pi/g]}.$$

Inefficiency Factors for Concentrated IS Density



Large sample inefficiency factors $\text{InEff}_\infty = \Omega(h)/\mathbb{V}_\pi[h]$ (dashed) and as their small sample approximations (solid) based on $N_{run} = 1,000$. We consider $h(\theta) = \theta$ (triangles) and $h(\theta) = \theta^2$ (circles). The solid line (no symbols) depicts the approximate inefficiency factor $1 + \mathbb{V}_g[\pi/g]$.

Inefficiency Factors for Diffuse IS Density



Large sample inefficiency factors $\text{InEff}_\infty = \Omega(h)/\mathbb{V}_\pi[h]$ (dashed) and as their small sample approximations (solid) based on $N_{run} = 1,000$. We consider $h(\theta) = \theta$ (triangles) and $h(\theta) = \theta^2$ (circles). The solid line (no symbols) depicts the approximate inefficiency factor $1 + \mathbb{V}_g[\pi/g]$.

Toward Frequentist Analysis: Likelihood Function

- Define $\hat{V}^{-1} = X'X = \sum y_{t-1}^2$
- In terms of reduced form parameter (note that $\hat{\phi}$ is sufficient statistic):
$$p(Y|\phi) \quad \propto \quad \exp\left\{ -\frac{1}{2}\hat{V}^{-1}(\phi - \hat{\phi})^2 \right\}$$
- Define auxiliary parameter $\alpha = \theta - \phi \in [0, 1]$.
- In terms of structural and auxiliary parameter:
$$p(Y|\theta, \alpha) \quad \propto \quad \exp\left\{ -\frac{1}{2}\hat{V}^{-1}(\theta - \alpha - \hat{\phi})^2 \right\}$$
- α is a nuisance parameter with bounded domain.

Frequentist Analysis

- Problem: we have a non-identifiable nuisance parameter α in the objective function.
- Concentrate out nuisance parameter?
- Integrate out nuisance parameter?

Concentrate Out Nuisance Parameter

- Concentrate out nuisance parameter α.

$$\alpha_*(\theta) = \text{argmin}_{\alpha \in \mathcal{A}_\theta} \; \hat{V}^{-1}(\theta - \alpha - \hat{\phi})^2$$

- Profile likelihood function

$$p(Y|\theta, \alpha_*(\theta)) \propto \exp\left\{ -\frac{1}{2} Q(\theta|\hat{\phi}) \right\}$$

where

$$Q(\theta|\hat{\phi}) = \begin{cases} \hat{V}^{-1}(\hat{\phi} - \theta)^2 & \text{if } \theta \leq \hat{\phi} \\ 0 & \text{if } \theta \in \Theta(\hat{\phi}) \\ \hat{V}^{-1}(\hat{\phi} - \theta + \lambda)^2 & \text{if } \hat{\phi} + \lambda \leq \theta \end{cases}$$

Frequentist Analysis of Profile LH

- Approximate distribution of profile objective function

$$Q(\theta|\hat{\phi}) \overset{\text{approx}}{\sim} \begin{cases} (\mathcal{Z}-s)^2 & \text{if } \mathcal{Z} \geq s \\ 0 & \text{otherwise} \\ (\mathcal{Z}-s+\hat{V}^{-1/2}\lambda)^2 & \text{if } \mathcal{Z} \leq s-\hat{V}^{-1/2}\lambda \end{cases},$$

  where $\mathcal{Z} \sim N(0,1)$ and $s = \hat{V}^{-1/2}(\theta-\phi)$.

- Approximation relies on a CLT that implies

$$\hat{V}^{-1/2}(\hat{\phi}-\phi) \Longrightarrow N(0,1).$$

- Note that $\hat{V}^{-1} = (X'X)^{-1/2}$ and expands at rate $T^{1/2}$.

Frequentist Analysis of Profile LH

- A $1-\tau$ frequentist confidence interval for θ in the identified set $\Theta(\phi)$ can be constructed as follows:

$$CS_F^{\theta}(\hat{\phi}) = \left\{ \theta \;\middle|\; Q(\theta|\hat{\phi}) \leq c_T^2(\theta) \right\}.$$

- The sequence of critical values or threshold levels c_T satisfies the constraint

$$\lim_{T \longrightarrow \infty} \inf_{\phi \in \Phi} \inf_{\theta \in \Theta(\phi)} P_{\phi}^{\hat{\phi}_T} \{ \, Q_T(\theta|\hat{\phi}_T) \leq c_T^2(\theta) \} \geq 1-\tau.$$

- Approximately, c_T has to solve the following equation:

$$\Phi_N(\hat{V}^{-1/2}\lambda + c_T) - \Phi_N(-c_T) = 1-\tau.$$

- Here critical value is independent of θ and generates a *level set*. If $\hat{V}^{-1/2}\lambda$ is large 95% interval is:

$$CS_F^{\theta}(Y^n) = \left[\hat{\phi}_n - 1.64\hat{V}^{1/2} \,,\; \hat{\phi}_n + \lambda + 1.64\hat{V}^{1/2} \right]$$

Remarks

- If $\lambda = 0$, $CS_F^\theta(Y)$ and $CS_B^\theta(Y)$ are identical.
- If $\lambda > 0$, then frequentist and Bayesian intervals are different:
  - $CS_F^\theta(Y)$ expands the boundaries of the interval $\Theta(\hat{\phi})$ by approximately $\Phi^{-1}(1-\tau)\hat{V}^{1/2}$.
  - $CS_B^\theta(Y)$ lies strictly in the interior of $\Theta(\hat{\phi})$ for large T.
- Frequentist inference is non-standard because
  - nuisance parameter problem (concentrate or integrate!);
  - sampling distribution of $Q(\theta|\phi)$ changes with distance to boundary of $\Theta(\phi_0)$;
- Literature also considers confidence sets for $\Theta(\phi_0)$.