Nonstandard Inference

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

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

- 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 $\mathbb{V}[u_1] = 1$):

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

- 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.

Unit Roots

- The equivalence between sampling and posterior distributions breaks drown when $\phi_0 = 1$ (unit root) and

$$y_t = y_{t-1} + u_t.$$

- The derivation of the posterior distribution remains correct:

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

with

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

Note that there is no need to approximate the large sample behavior of $X'X$ and $X'Y$ to implement Bayesian inference.

- The frequentist approximation breaks down:

$$\sqrt{T}(\hat{\phi} - \phi_0) \not\Longrightarrow N\big(0, \mathbb{E}[y_{t-1}^2]\big).$$

Sums of Random Variables

- Assume that $\phi_0 = 1$ and $y_0 = 0$.
- Thus, the process y_t can be represented as

$$y_t = \sum_{\tau=1}^{t} u_\tau$$

- Summations will range from $t = 1$ to T unless stated otherwise. The central limit theorem for *iid* random variables implies

$$\frac{y_T}{\sqrt{T}} = \frac{1}{\sqrt{T}} \sum_{\tau=1}^{T} u_\tau \Longrightarrow N(0,1)$$

This suggests that

$$\frac{1}{T} \sum y_t = \frac{1}{\sqrt{T}} \sum \left[\sqrt{\frac{t}{T}} \frac{1}{\sqrt{t}} \sum_{\tau=1}^{t} u_\tau \right]$$

will not converge to a constant in probability but instead to a random variable.

A Functional Central Limit Theorem

- Let $\mathcal{C}$ be the space of continuous functions on the interval $[0, 1]$.
- We will define a probability distribution for the function space $\mathcal{C}$. This probability distribution is called "Wiener measure".
- Whenever we draw an element from the probability space we obtain a function $W(s)$, $s \in [0, 1]$, which properties described on the following two slides.

A Functional Central Limit Theorem

- If we repeatedly draw functions under the Wiener measure and evaluate these functions at a particular value $s = s'$, then

$$\mathbb{P}[\{W(s') \leq w\}] = \frac{1}{\sqrt{2\pi s'}} \int_{-\infty}^{w} e^{-u^2/2s'} du$$

that is,

$$W(s') \sim N(0, s')$$

If $s' = 0$ then the equations is interpreted to mean $\mathbb{P}[\{W(0) = 0\}] = 1$. Thus $W(0) = 0$ with probability one.

A Functional Central Limit Theorem

- The random function $W(s)$ has independent increments. If

  $$0 \leq s_1 \leq s_2 \leq \ldots \leq s_k \leq 1$$

  Then the random variables

  $$W(s_2) - W(s_1),\ W(s_3) - W(s_2)\ \ldots,\ W(s_k) - W(s_{k-1})$$

  are independent.
- The random function $W(s)$ is continuous on $s \in [0, 1]$. Otherwise we would contradict our assumption that we are sampling from the space of continuous functions.

A Functional Limit Theorem

- We will convert the discrete trajectories of the random walk into continuous random functions on the interval $s \in [0,1]$.
- Define the partial sum process

$$Y_T(s) = \frac{1}{\sqrt{T}} \sum \{t \leq \lfloor Ts \rfloor\} u_t$$

  where $\lfloor x \rfloor$ denotes the integer part of x.
- Since we assumed that $u_t \sim iid(0,1)$, the partial sum process is a random step function.
- Step functions are not continuous, but it is easy to interpolate the steps as follows.

$$\bar{Y}_T(s) = \frac{1}{\sqrt{T}} \sum \{t \leq \lfloor Ts \rfloor\} u_t + (Ts - \lfloor Ts \rfloor) u_{\lfloor Ts \rfloor + 1} / \sqrt{T}$$

A Functional Limit Theorem

We now have two random experiments to generate continuous functions:

1. Draw a function $W(s)$ from the Wiener distribution. We did not examine how to do the sampling in practice, but since the Wiener distribution is well-defined, it is theoretically possible.
2. Generate a sequence $u_1, \ldots, u_T$, where $u_t \sim iid(0,1)$ and compute $\bar{Y}_T(s)$.

It turns out that the two experiments become very similar as T in the second experiment tends to infinity.

Donsker's Functional Central Limit Theorem

Let $u_t \sim iid(0,\sigma^2)$. Then

$$Y_T(s) = \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^{T} \{t \leq \lfloor Ts \rfloor\} u_t \Longrightarrow W(s) \quad \square$$

A Functional Central Limit Theorem

- The functional central limit theorem is a statement about convergence of probability distributions on function spaces.
- The proof is quite complicated and requires a rigorous definition of convergence of probability measures on metric spaces.
- First, one has to establish the convergence of finite-dimensional distributions

$$\left[Y_T(s_1), \ldots, Y_T(s_k) \right] \Longrightarrow \left[W(s_1), \ldots, W(s_k) \right]$$

for arbitrary finite-dimensional vectors $[s_1, \ldots, s_k]$.

A Functional Central Limit Theorem

- It is relatively straightforward to verify that for any fixed $s = s_*$, $Y_T(s_*) \Longrightarrow W(s_*)$, that is, $Y_T(s_*) \Longrightarrow N(0, s_*)$:
- Verify that $\bar{Y}_T(s_*) - Y_T(s_*) \stackrel{p}{\longrightarrow} 0$.
- Define $T_* = \lfloor Ts \rfloor$. Use the CLT for *iid* random variables to verify that

$$
\begin{aligned}
Y_T(s_*) &= \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^{T} \{t \leq \lfloor Ts_* \rfloor\} u_t \\
&= \sqrt{\frac{T_*}{T}} \frac{1}{\sigma\sqrt{T_*}} \sum_{t=1}^{T_*} u_t \Longrightarrow \sqrt{s} N(0,1) \quad \square
\end{aligned}
$$

A Functional Central Limit Theorem

- Second, one has to establish *tightness*.
- A family $\mathcal{P}$ of probability measures is *tight* if there exists a compact set K for which $P[K] > 1 - \epsilon$ for all $P \in \mathcal{P}$ for some $\epsilon > 0$.
- Counterexample: $F_n(x) = 0$ for $x \leq n$ and $F_n(x) = 1$ otherwise.
- Tightness is established by ensuring that partial sums do not fluctuate too much. For every $\epsilon, \eta > 0$ there exists a $\delta > 0$ such that

$$P\left\{ \sup_{|r-s|<\delta} |Y_T(r) - Y_T(s)| \quad > \epsilon \right\} < \eta.$$

Continuous Mappings

- **Fact**: If h is a measurable mapping of a metric space $\mathcal{S}$ into another metric space $\mathcal{S}'$, then each probability distribution P on $\mathcal{S}$ induces on $\mathcal{S}'$ a unique probability distribution hP. Suppose that h is a continuous mapping. If the sequence of distributions $P_T \Longrightarrow P$, then $hP_T \Longrightarrow hP$. $\square$
- We are interested in the metric space $\mathcal{S} = \mathcal{C}$ and mappings of the form

$$h : \mathcal{C} \mapsto \mathcal{C} \qquad h(f(s)) = f(s)^2$$

$$h : \mathcal{C} \mapsto \mathbb{R} \qquad h(f(s)) = \int_0^1 f(s)ds$$

$$h : \mathcal{C} \mapsto \mathbb{R} \qquad h(f(s)) = \int_0^1 f(s)^2 ds$$

Continuous Mappings

- Consider the expression

$$\begin{aligned} \frac{1}{T^{3/2}} \sum y_t &= \frac{1}{T} \sum \left[\frac{1}{\sqrt{T}} \sum_{\tau=1}^{T} \{\tau \leq t\} u_\tau \right] \\ &= \frac{1}{T} \sum_{t=1}^{T-1} Y_T(t/T) + \frac{1}{T} Y_T(1) \\ &= \int_0^1 Y_T(s) ds + \text{small} \end{aligned}$$

- Similarly,

$$\frac{1}{T^2} \sum y_t^2 = \frac{1}{T} \sum Y_T(t/T)^2 = \int_0^1 Y_T^2(s) ds + \text{small}$$

Continuous Mappings

Result: Suppose that $y_t = y_{t-1} + u_t$, where $u_t \sim iid(0,1)$ and $y_0 = 0$. Then

$$\frac{1}{T^{3/2}} \sum y_t \Longrightarrow \int W(s)ds$$

$$\frac{1}{T^2} \sum y_t^2 \Longrightarrow \int W(s)^2 ds$$

where $W(s)$ denotes a standard Wiener process.

Convergence to a Stochastic Integral

- To be able to analyze regression models we have to find the limiting behavior for expressions such as

$$\frac{1}{T}\sum y_{t-1}u_t.$$

- Roughly speaking, one can regard $u_t/\sqrt{T}$ as a stochastic increment $\Delta Y_T(r)$ and show that the sum converges to a stochastic integral.

- **Result:** Suppose that $y_t = y_{t-1} + u_t$, where $u_t \sim iid(0,1)$ and $y_0 = 0$. Then

$$\frac{1}{T}\sum y_{t-1}u_t \Longrightarrow \int W(s)dW(s),$$

where $W(s)$ denotes a standard Wiener process.

Main Result

If $\phi_0 = 1$ the OLS estimator has the following limit distribution

$$T(\hat{\phi} - 1) = \frac{\frac{1}{T}\sum y_{t-1}u_t}{\frac{1}{T^2}\sum y_{t-1}^2} \Longrightarrow \frac{\int_0^1 W(s)dW(s)}{\int_0^1 W(s)^2 ds}.$$

A Helicopter Tour

- This excursion is based on Sims and Uhlig (Econometrica, 1991).
- Assume that $\phi \sim \mathcal{U}[0.8, 1.1]$
- Data are generated by drawing a ϕ from the prior, setting $y_0 = 0$ and simulating

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

- The MLE/OLS estimator is given by

$$\hat{\phi} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2}.$$

- The likelihood function is

$$p(Y|\phi) \propto \exp\left\{ -\frac{1}{2}\left[(\phi - \hat{\phi})^2 \sum y_{t-1}^2 \right] \right\}.$$

A Helicopter Tour

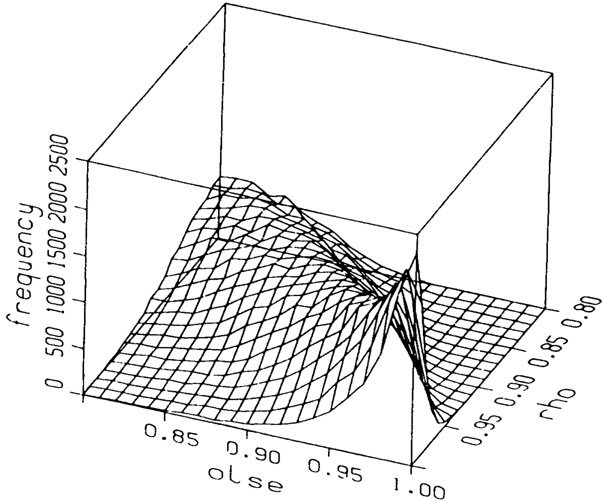

FIGURE 1.—Joint frequency distribution of $\hat{\rho}$ and ρ.

A Helicopter Tour

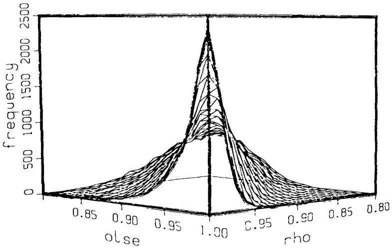

FIGURE 2.—Joint frequency distribution of $\hat{\rho}$ and ρ.

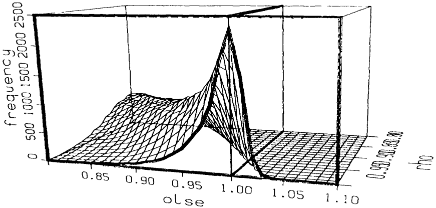

FIGURE 3.—Joint frequency distribution of $\hat{\rho}$ and ρ sliced along $\rho = 1$.

A Helicopter Tour

UNDERSTANDING UNIT ROOTERS

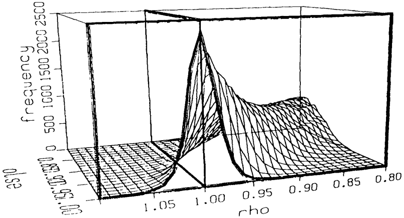

FIGURE 4.—Joint frequency distribution of $\hat{\rho}$ and ρ sliced along $\hat{\rho} = 1$.

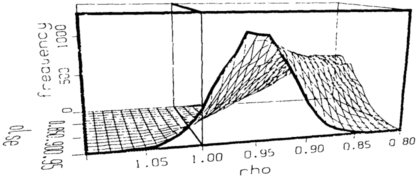

FIGURE 5.—Joint frequency distribution of $\hat{\rho}$ and ρ sliced along $\hat{\rho} = .95$.

Summary

- In non-stationary / unit-root models frequentist asymptotics work quite differently from "regular" stationary models.
- It is much more complicated to construct frequentist tests and confidence intervals.
- The numerical equivalence between Bayesian and frequentist confidence intervals in large samples breaks down.