Midterm Exam, Econ 510 **YOUR NAME:**
Monday 2/23/26, 10:00-11:30am
Patrik Guggenberger, Department of Economics, Penn State

Answer all four questions for a total of 40 points (each question is worth 10 points). Closed book exam, no calculators or cell phones allowed. You can write on the front and back sides of each page. GOOD LUCK!

1. Consider the model $Y_i = \alpha_0 + \beta_0 X_i + U_i$ for $i = 1, ..., n$. Let Z_i be another random variable (that we will call an *instrumental variable*). Suppose (Y_i, X_i, U_i, Z_i) are i.i.d. random variables with finite means and variances. The instrumental variables estimator of β_0 is defined by

$$\widehat{\beta}_{IV,n} = \left(n^{-1} \sum_{i=1}^{n} (Z_i - \overline{Z}_n) X_i \right)^{-1} n^{-1} \sum_{i=1}^{n} (Z_i - \overline{Z}_n) Y_i,$$

where $\overline{Z}_n$ denotes the sample average of the Z_i. Do not use high level theorems (like Theorem 12.1... from the lecture notes) here. Derive the limiting results from scratch.

a) Find the probability limit of $\widehat{\beta}_{IV,n}$ under the weakest additional conditions that you can (and state what these conditions are). **2 points**

b) Under what additional conditions is $\widehat{\beta}_{IV,n}$ consistent for the true value β_0? **1 point**

c) Find the asymptotic distribution of $n^{1/2}(\widehat{\beta}_{IV,n} - \beta_0)$ under the conditions above and the additional conditions that you specify in parts a. and b. and any additional conditions that are needed. **2 points**

d) Provide an estimator for the variance of the limiting distribution in c. Show that the proposed estimator is consistent. **1 point**

e) Write down the Wald statistic to test the null hypothesis $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0$. Describe the intuition underlying the construction of that statistic. **2 points**

f) Explain how you would construct an (asymptotically valid) 95% confidence region for β based on a Wald test. Provide a discussion about the precise sense of asymptotic validity used here. **2 points**

1. (a) **Solutions:** a.

$$\begin{aligned}
\widehat{\beta}_{IV,n} &= \left(n^{-1}\sum_{i=1}^{n} Z_iX_i - \overline{Z}_n\overline{X}_n\right)^{-1}(n^{-1}\sum_{i=1}^{n} Z_iY_i - \overline{Z}_n\overline{Y}_n) \\
&\to_p (EZ_iX_i - EZ_iEX_i)^{-1}(EZ_iY_i - EZ_iEY_i) \\
&= (EZ_iX_i - EZ_iEX_i)^{-1}[EZ_i(\beta_0X_i + U_i) - EZ_iE(\beta_0X_i + U_i)] \\
&= \beta_0 + (EZ_iX_i - EZ_iEX_i)^{-1}[EZ_iU_i - EZ_iEU_i],
\end{aligned}$$

where the first equality follows by rearranging terms, the convergence in probability statement follows by the WLLN and Slutsky assuming $EZ_iX_i - EZ_iEX_i \neq 0$ (instrument relevance), $E||Z_iX_i|| < \infty$, $E||Z_iY_i|| < \infty$, $E||Z_i|| < \infty$, $E||X_i|| < \infty$, and $E||Y_i|| < \infty$, the second equality follows by plugging in the definition of Y_i, and the last equality follows from simple rearrangement.
b. If $cov(Z_i, U_i) = EZ_iU_i - EZ_iEU_i = 0$ (instrument exogeneity) then $\widehat{\beta}_{IV,n} \to_p \beta_0$.
c. Note that by plugging in the definition of Y_i and simple rearrangements we obtain

$$n^{1/2}(\widehat{\beta}_{IV,n}-\beta_0) = \left(n^{-1}\sum_{i=1}^{n} Z_iX_i - \overline{Z}_n\overline{X}_n\right)^{-1} n^{1/2}(n^{-1}\sum_{i=1}^{n} Z_iU_i-\overline{Z}_n\overline{U}_n).$$

By a. $\left(n^{-1}\sum_{i=1}^{n} Z_iX_i - \overline{Z}_n\overline{X}_n\right)^{-1} \to_p (EZ_iX_i - EZ_iEX_i)^{-1}$. Assume $EZ_iU_i = EU_i = EZ_i = 0$, $E||Z_iU_i||^2 < \infty, E||Z_i||^2 < \infty$, $E||U_i||^2 < \infty$, and $E(Z_iU_i)^2 > 0$. Then by the CLT we obtain

$$n^{-1/2}\sum_{i=1}^{n} Z_iU_i - (n^{-1/2}\sum_{i=1}^{n} Z_i)(n^{-1}\sum_{i=1}^{n} U_i) \to_d N(0, E(Z_iU_i)^2)$$

because $(n^{-1/2}\sum_{i=1}^{n} Z_i)(n^{-1}\sum_{i=1}^{n} U_i) = O_p(1)o_p(1) = o_p(1)$. Therefore,

$$n^{1/2}(\widehat{\beta}_{IV,n} - \beta_0) \to_d N(0, E(Z_iU_i)^2/(EZ_iX_i)^2).$$

d. We estimate $E(Z_iU_i)^2/(EZ_iX_i)^2$ by a sample counterpart

$$\widehat{V}_n = n^{-1}\sum_{i=1}^{n}(Z_i\widehat{U}_i)^2/(n^{-1}\sum_{i=1}^{n}(Z_iX_i))^2$$

where $\widehat{U}_i := Y_i - \widehat{\alpha}_0 + \widehat{\beta}_0X_i$ and $(\widehat{\alpha}_0, \widehat{\beta}_0)$ can be taken as the IV estimator of (α_0, β_0). Consistency then follows by the usual steps.
e. Using the notation from Lecture 14,$h(\alpha, \beta) = \beta$ and $H = (0, 1)$. The Wald statistic is given by

$$W_n = \frac{n\widehat{\beta}^2_{IV,n}}{\widehat{V}_n}.$$

The intuition is that if the null hypothesis is true then $\widehat{\beta}_{IV,n}$ should be close to zero and therefore W_n (a quadratic form in $\widehat{\beta}_{IV,n}$) should be small. In contrast, if the null is false then $\widehat{\beta}_{IV,n}$ should be different from zero and $n\widehat{\beta}^2_{IV,n}$ and thus the Wald statistic should blow up. Therefore the Wald test rejects if W_n is "too big".

f. One can obtain a (pointwise asymptotically valid) CI with coverage probability $1-\alpha$ by inverting the Wald test of $H_0 : \beta_0 = \beta$ vs $H_1 : \beta_0 \neq \beta$ at nominal size α. To test that hypothesis the Wald statistic reads $W_n(\beta_0 = \beta) = \frac{n(\widehat{\beta}_{IV,n}-\beta)^2}{\widehat{V}_n}$That is,

$$CI_n(\beta_0) = \{\beta \in R; W_n(\beta_0 = \beta) \leq \chi^2_{1,1-\alpha}\},$$

where $\chi^2_{1,1-\alpha}$ denotes the $1-\alpha$ quantile from a chi-square with one degrees of freedom. Here we would take $\alpha = 5\%$. Then the probability that $CI_n(\beta_0)$ contains the true β_0 - for a fixed DGP - converges as n goes to infinity to $1-\alpha$. However, as we discussed in class, that statement is not true if one allows for the DGP to change with the sample size. For any sample size, one can find DGPs (like a DGP where the IV has very small correlation with X_i) under which the coverage probability is much smaller than $1-\alpha$..

2. a) Under Assumptions EE3 and CF-NS (stated below) we derived the limiting distribution of the GMM estimator $\widehat{\theta}_n$ with nonsmooth stochastic criterion function, namely

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \rightarrow_d N(0, (\Gamma'\Gamma)^{-1}\Gamma' V_0 \Gamma(\Gamma'\Gamma)^{-1}).$$

Provide estimators for V_0 and Γ and discuss their consistency under appropriate conditions. **4 points**

b) Suppose $X_n \rightarrow_p \mu$ for some constant μ. Show that there exists a sequence of positive constants δ_n such that $\delta_n \rightarrow 0$ as $n \rightarrow \infty$ and $P(|X_n - \mu| > \delta_n) \rightarrow 0$ as $n \rightarrow \infty$. **6 points**

Setup and Assumptions for the GMM case with nonsmooth stochastic criterion function:

Let $Q_n(\theta) = ||\overline{g}_n(\theta)||$, where $\overline{g}_n(\theta) = n^{-1}\sum_{i=1}^n g(W_i, \theta)$, and $g(\theta) = Eg(W_i, \theta)$.

Assumption CF-NS: (i) θ_0 *is in the interior of* Θ.

(ii) $g(\theta)$ *is differentiable at* θ_0 *with* $\Gamma = (\partial/\partial\theta')g(\theta_0)$ of full rank $d \leq k$.

(iii) $g(\theta_0) = 0$.

(iii) $\sqrt{n}\overline{g}_n(\theta_0) \rightarrow_d N(0, V_0)$.

(iv) For every sequence of positive constants $\{\delta_n\}_{n\geq 1}$ that converges to zero,

$$\sup_{\theta\in\Theta, ||\theta-\theta_0||<\delta_n} \sqrt{n}\, ||\overline{g}_n(\theta) - g(\theta) - \overline{g}_n(\theta_0)|| \rightarrow_p 0.$$

Assumption EE3: (i) $Q_n(\widehat{\theta}_n) = \inf_{\theta\in\Theta} Q_n(\theta) + o_p(n^{-1/2})$ and (ii) $\widehat{\theta}_n \rightarrow_p \theta_0$.

Solutions:

a) This is taken straight from the lecture notes. See lecture notes 16, (16.24)-(16.26).

b) See homework 4. We proceed by contradiction. If the statement was false then for all sequences δ_n such that $\delta_n \to 0$ there exists an $\varepsilon^* > 0$ such that for all $N < \infty$ there exists an $n \geq N$ such that

$$P(|X_n - \mu| > \delta_n) > \varepsilon^*. \tag{1}$$

Given that $X_n \to_p \mu$ for every $m \in N$ there is an $N_m < \infty$ such that for all $n \geq N_m$

$$P(|X_n - \mu| > m^{-1}) < \varepsilon^*.$$

Of course, N_m can be chosen to be increasing. Define a sequence δ_n as m^{-1} for $N_m \leq n < N_{m+1}$ and equal to 1 for $1 \leq n < N_1$. Then $\delta_n \to 0$ and $P(|X_n - \mu| > \delta_n) < \varepsilon^*$ for all $n \geq N_1$ which contradicts (1).

3. Consider a moment condition model $Eg(W_i, \theta) = 0 \in R^k$ for $\theta = \theta_0 \in R^p$ where $k \geq p$. Assume CF and EE2 (the high level assumptions that imply asymptotic normality of EE) are satisfied for the GMM model, *except that* $\Gamma = E(\partial/\partial\theta) g(W_i, \theta_0) \in R^{k\times p}$ has rank $p-1$ with its p-th column being a zero vector. In particular, assume the GMM estimator $\widehat{\theta}_n = (\widehat{\theta}'_{\cdot n}, \widehat{\theta}_{pn})'$ (for simplicity with weighting matrix $A_n = I_k$) is consistent, where $\widehat{\theta}'_{\cdot n}$ denotes the first $p-1$ and $\widehat{\theta}_{pn}$ the p-th component of $\widehat{\theta}_n$, respectively. We will establish that $\widehat{\theta}_{pn} - \theta_{p0} = O_p(n^{-1/4})$, where $\theta_0 = (\theta'_{\cdot 0}, \theta_{p0})'$. Let $\widehat{g}_n(\theta) = \sum_{i=1}^n g(W_i, \theta)/n$. Define[1]

$$\widehat{g}_{\cdot n}(\theta_{\cdot}) = \widehat{g}_n(\theta_{\cdot}, \widehat{\theta}_{pn}) \text{ and } \widehat{g}_{pn}(\theta_p) = \widehat{g}_n(\theta_{\cdot 0}, \theta_p).$$

i) Do a mean value expansion of $\widehat{g}_{\cdot n}(\widehat{\theta}_{\cdot n})$ about $\theta_{\cdot 0}$ and a second order Taylor expansion of $\widehat{g}_{pn}(\widehat{\theta}_{pn})$ about θ_{p0}. Combine the two expansions and, using consistency, establish that

$$\widehat{\theta}_{\cdot n} - \theta_{\cdot 0} = (\widehat{\Gamma}'_{\cdot n}\widehat{\Gamma}_{\cdot n})^{-1}\widehat{\Gamma}'_{\cdot n}[(\widehat{g}_n(\widehat{\theta}_n) - \widehat{g}_n(\theta_0)) - \frac{1}{2}\widehat{B}_{pn}(\widehat{\theta}_{pn} - \theta_{p0})^2] + o_p(n^{1/2}), \quad (2)$$

where $\widehat{\Gamma}_{\cdot n} = (\partial/\partial\theta_{\cdot})\widehat{g}_n(\overline{\theta}_{\cdot n}, \widehat{\theta}_{pn})$ for some vector $\overline{\theta}_{\cdot n}$ on the line segment joining $\theta_{\cdot 0}$ and $\widehat{\theta}_{\cdot n}$ and $\widehat{B}_{pn} = (\partial^2/\partial\theta_p^2)\widehat{g}_n(\theta_{\cdot 0}, \overline{\theta}_{pn})$ for some $\overline{\theta}_{pn}$ between θ_{p0} and $\widehat{\theta}_{pn}$. **5 points**

ii) Conclude that

$$\widehat{g}_n(\widehat{\theta}_n) = \widehat{g}_n(\theta_0) + P_{\widehat{\Gamma}_{\cdot n}}(\widehat{g}_n(\widehat{\theta}_n) - \widehat{g}_n(\theta_0)) + \frac{1}{2}M_{\widehat{\Gamma}_{\cdot n}}\widehat{B}_{pn}(\widehat{\theta}_{pn} - \theta_{p0})^2 + o_p(n^{-1/2}),$$

where $P_A = A(A'A)^{-1}A'$ and $M_A = I_k - P_A$ for any matrix A with k rows. **1 point**

iii) Use ii) and $||\widehat{g}_n(\widehat{\theta}_n)||^2 \leq ||\widehat{g}_n(\theta_0)||^2$ to conclude that

$$n^{1/2}(\widehat{\theta}_{pn} - \theta_{p0})^2 O_p(1) + (\frac{1}{4}B'M_\Gamma B + o_p(1))n(\widehat{\theta}_{pn} - \theta_{p0})^4 = O_p(1),$$

where M_Γ and B denote the probability limits of $M_{\widehat{\Gamma}_{\cdot n}}$ and $\widehat{B}_{pn}$, respectively. **2 points**

iv) Assume that $B'M_\Gamma B$ is not the zero matrix. Conclude from iii) that $\widehat{\theta}_{pn} - \theta_{p0} = O_p(n^{-1/4})$. Show that that implies that $\widehat{\theta}_{\cdot n} - \theta_{\cdot 0} = O_p(n^{-1/2})$. **2 points**

---

[1] We leave out transpose signs for the arguments of $\widehat{g}_n$; e.g. we write $(\theta_{\cdot}, \widehat{\theta}_{pn})$ rather than $(\theta'_{\cdot}, \widehat{\theta}_{pn})'$ to simplify notation.

Solution

i) By a mean value and a second order Taylor expansion we obtain

$$\widehat{g}_n(\widehat{\theta}_{\cdot n}, \widehat{\theta}_{pn}) = \widehat{g}_n(\theta_{\cdot 0}, \widehat{\theta}_{pn}) + (\partial/\partial\theta_{\cdot})\widehat{g}_n(\overline{\theta}_{\cdot n}, \widehat{\theta}_{pn})(\widehat{\theta}_{\cdot n} - \theta_{\cdot 0})$$

and

$$\widehat{g}_n(\theta_{\cdot 0}, \widehat{\theta}_{pn}) = \widehat{g}_n(\theta_{\cdot 0}, \theta_{p0}) + (\partial/\partial\theta_p)\widehat{g}_n(\theta_{\cdot 0}, \theta_{p0})(\widehat{\theta}_{pn} - \theta_{p0}) + \frac{1}{2}(\partial^2/\partial\theta_p^2)\widehat{g}_n(\theta_{\cdot 0}, \overline{\theta}_{pn})(\widehat{\theta}_{pn} - \theta_{p0})^2.$$

Plugging the second expression into the first one we get

$$\begin{aligned} & \widehat{g}_n(\widehat{\theta}_n) \\ = \; & \widehat{g}_n(\theta_{\cdot 0}, \theta_{p0}) + (\partial/\partial\theta_p)\widehat{g}_n(\theta_0)(\widehat{\theta}_{pn} - \theta_{p0}) + (\partial/\partial\theta_{\cdot})\widehat{g}_n(\overline{\theta}_{\cdot n}, \widehat{\theta}_{pn})(\widehat{\theta}_{\cdot n} - \theta_{\cdot 0}) + \frac{1}{2}\widehat{B}_{pn}(\widehat{\theta}_{pn} - \theta_{p0})^2 \\ = \; & \widehat{g}_n(\theta_0) + \widehat{\Gamma}_{\cdot n}(\widehat{\theta}_{\cdot n} - \theta_{\cdot 0}) + \frac{1}{2}\widehat{B}_{pn}(\widehat{\theta}_{pn} - \theta_{p0})^2 + o_p(n^{-1/2}), \end{aligned} \tag{3}$$

where for the last equality we use $(\partial/\partial\theta_p)\widehat{g}_n(\theta_0)(\widehat{\theta}_{pn} - \theta_{p0}) = O_p(n^{-1/2})o_p(1) = o_p(n^{-1/2})$ (where in the first equality we use consistency of $\widehat{\theta}_{pn}$ and $(\partial/\partial\theta_p)\widehat{g}_n(\theta_0) = O_p(n^{-1/2})$ which holds by a CLT because by assumption $E(\partial/\partial\theta_p)g(W_i, \theta_0)$ equals 0). Finally, premultiply by $(\widehat{\Gamma}'_{\cdot n}\widehat{\Gamma}_{\cdot n})^{-1}\widehat{\Gamma}'_{\cdot n}$ and solve for $\widehat{\theta}_{\cdot n} - \theta_{\cdot 0}$.

ii) That follows straightforwardly by plugging in the result from (2) into (3).

iii) From ii) and carefully assessing the order of each term, we obtain that

$$||\widehat{g}_n(\widehat{\theta}_n)||^2 = ||\widehat{g}_n(\theta_0)||^2 + (\widehat{\theta}_{pn} - \theta_{p0})^2 O_p(n^{-1/2}) + \frac{1}{4}\widehat{B}'_{pn}M_{\widehat{\Gamma}_{\cdot n}}\widehat{B}_{pn}(\widehat{\theta}_{pn} - \theta_{p0})^4 + O_p(n^{-1})$$

noting that $\widehat{g}_n(\widehat{\theta}_n) - \widehat{g}_n(\theta_0) = O_p(n^{-1/2})$. The desired result therefore is implied by $||\widehat{g}_n(\widehat{\theta}_n)||^2 \leq ||\widehat{g}_n(\theta_0)||^2$.

iv) If not $\widehat{\theta}_{pn} - \theta_{p0} = O_p(n^{-1/4})$, then $n(\widehat{\theta}_{pn} - \theta_{p0})^4$ is not $O_p(1)$ and would diverge faster than $n^{1/2}(\widehat{\theta}_{pn} - \theta_{p0})^2$. But then, the left hand side in (iii) could not be $O_p(1)$.

By i) and $n(\widehat{\theta}_{pn} - \theta_{p0})^4$ we get the desired result $\widehat{\theta}_{\cdot n} - \theta_{\cdot 0} = O_p(n^{-1/2})$.

4. Assume $(X_i, Y_i) \in R^2$, $i = 1, ..., n$, are i.i.d. with distribution F (that may depend on n). Let $\mu_X = E_F X_i$ and $\mu_Y = E_F Y_i$. We are interested in a CI C_n for the parameter

$$\theta = \max\{\mu_X, \mu_Y\}$$

(that may depend on n) that satisfies

$$AsyCS = \lim \inf_{n\to\infty} \inf_{F\in\mathcal{F}} P_F(\theta \in C_n) \geq 1 - \alpha,$$

i.e. the asymptotic confidence size is nonsmaller than the nominal coverage probability, where $P_F(\cdot)$ denotes probability when the true distribution is F and the set of distributions $\mathcal{F}$ is suitably restricted such that WLLNs and CLTs can be applied. Namely, take

$$\begin{aligned}\mathcal{F} = \mathcal{F}_{\delta,M} = \{F \quad : \quad & F \text{ is the distribution of a random vector } (X, Y);\\ & E_F||(X,Y)||^{2+\delta} \leq M,\ \delta \leq \sigma_X^2, \sigma_Y^2 \leq M \text{ and } |\rho| \leq 1 - \delta\},\end{aligned}$$

where ρ, σ_X^2, and σ_Y^2 denote the correlation between X and Y and the variances of these variables and $M < \infty$ and $\delta > 0$.

Denote by $\overline{X}_n$ and $\overline{Y}_n$ the sample averages and by $\widehat{\sigma}_{Xn}^2$ and $\widehat{\sigma}_{Yn}^2$ the sample variances of the X_is and Y_is, respectively,

$$\overline{X}_n = n^{-1}\textstyle\sum_{i=1}^n X_i,\ \widehat{\sigma}_{Xn}^2 = n^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2...$$

Denote by

$$\begin{aligned} C_{\mu_X n} &= [\overline{X}_n - n^{-1/2}\widehat{\sigma}_{Xn} z_{1-\alpha/2}, \overline{X}_n + n^{-1/2}\widehat{\sigma}_{Xn} z_{1-\alpha/2}] \text{ and} \\ C_{\mu_Y n} &= [\overline{Y}_n - n^{-1/2}\widehat{\sigma}_{Yn} z_{1-\alpha/2}, \overline{Y}_n + n^{-1/2}\widehat{\sigma}_{Yn} z_{1-\alpha/2}] \end{aligned}$$

the standard $1-\alpha$ two-sided CI for μ_X and μ_Y, respectively, based on inverting a t-test.

Consider the CI C_n for θ defined as

$$C_{\mu_X n} \text{ if } \overline{X}_n \geq \overline{Y}_n \text{ and defined as } C_{\mu_Y n} \text{ otherwise.}$$

i) Discuss what happens to the coverage probability as $n \to \infty$ for fixed F. **4 points**

ii) Next define drifting sequences $\gamma_{nh} = (\gamma_{1n}, \gamma_{2n}, \gamma_{3n})$, where $\gamma_{1n} = \mu_{1n} - \mu_{2n}$, $\gamma_{2n} = (\sigma_{Xn}, \sigma_{Yn}, \rho_n)$, $\gamma_{3n} = F_n$, and

$$n^{1/2}\gamma_{1n} \to h_1,\ \gamma_{2n} \to h_2 = (h_{21}, h_{22}, h_{23})$$

for some $h_1 \in R \cup \{\pm\infty\}$ and $h_2 \in [\sqrt{\delta}, \sqrt{M}]^2 \times [\delta - 1, 1 - \delta]$. Derive the limiting coverage probability under γ_{nh}. **4 points**

iii) Using the result in ii) find a simple formula for $AsyCS$. **2 points**

1. **Solutions:** I am presenting a very lengthy solution with a lot of repetition.

   Note that by a Lyapounov triangular array CLT for row-wise i.i.d. random variables, if $\rho_n \to h_2$

   $$n^{1/2}((\overline{X}_n - \mu_{Xn})/\sigma_{Xn}, (\overline{Y}_n - \mu_{Yn})/\sigma_{Yn}) \to_d (\xi_{X,h_2}, \xi_{Y,h_2}) \sim N(0, \Sigma), \tag{4}$$

   where Σ has diagonal elements 1 and off diagonal element h_2. Note also that $\mu_{Xn} \in C_{\mu_X n}$ iff $n^{1/2}|\overline{X}_n - \mu_{Xn}|/\sigma_{Xn} \leq z_{1-\alpha/2}$ and likewise for $\mu_{Yn} \in C_{\mu_Y n}$.

   i) Assume first that μ_{Xn} and μ_{Yn} do not depend on n.

   If the means are not equal, wlog $\mu_X > \mu_Y$ say, then C_n has asymptotic coverage probability for $\theta = \mu_1$ equal to $1-\alpha$. This follows immediately, because by the law of large numbers $\overline{X}_n \to_p \mu_X$ and $\overline{Y}_n \to_p \mu_Y$ and therefore with probability approaching 1, $\overline{X}_n > \overline{Y}_n$. But $P(\mu_X \in C_{\mu_X n}) \to 1-\alpha$ by the central limit theorem (CLT).

   If the means are equal, $\mu_X = \mu_Y = \theta$, then $P(\theta \in C_n)$ converges to

   $$P(|\xi_{X,\rho}| < z_{1-\alpha/2} \text{ if } \sigma_X\xi_{X,\rho} \geq \sigma_Y\xi_{Y,\rho} \text{ or } |\xi_{Y,\rho}| < z_{1-\alpha/2} \text{ if } \sigma_X\xi_{X,\rho} < \sigma_Y\xi_{Y,\rho}), \tag{5}$$

   This follows from

   $$n^{1/2}((\overline{X}_{1n} - \mu_1)/\sigma_{1n}, (\overline{X}_{2n} - \mu_2)/\sigma_{2n}) \to_d (\xi_{1,\rho_{12}}, \xi_{2,\rho_{12}}). \tag{6}$$

   (ii) More generally, consider the case of drifting sequences $\gamma_{nh} = (\gamma_{1n}, \gamma_{2n}, \gamma_{3n})$, where $\gamma_{1n} = \mu_{1n} - \mu_{2n}$, $\gamma_{2n} = (\sigma_{Xn}, \sigma_{Yn}, \rho_n)$ and $\gamma_{3n} = F_n$. Consider sequences such that

   $$n^{1/2}\gamma_{1n} \to h_1, \ \gamma_{2n} \to h_2 = (h_{21}, h_{22}, h_{23}) \tag{7}$$

   for some $h_1 \in R_\infty$. If $h_1 \geq 0$, it follows that $P(\theta_n \in C_n(1-\alpha))$ converges to

   $$P(|\xi_{1,h_{23}}| < z_{1-\alpha/2} \text{ if } h_{21}\xi_{1,h_{23}} + h_1 \geq h_{22}\xi_{2,h_{23}}; |\xi_{2,h_{23}} - h_1/h_{22}| < z_{1-\alpha/2} \text{ if } h_{21}\xi_{1,h_{23}} + h_1 < h_{22}\xi_{2,h_{23}}) \tag{8}$$

   under assumptions on $\sigma^2_{Xn}, \sigma^2_{Yn}$, and ρ_n imposed in the parameter space. An analogous statement holds when $h_1 < 0$, namely $P(\theta_n \in C_n(1-\alpha))$ converges to

   $$P(|\xi_{1,h_{23}} + h_1/h_{21}| < z_{1-\alpha/2} \text{ if } h_{21}\xi_{1,h_{23}} + h_1 \geq h_{22}\xi_{2,h_{23}}; |\xi_{2,h_{23}}| < z_{1-\alpha/2} \text{ if } h_{21}\xi_{1,h_{23}} + h_1 < h_{22}\xi_{2,h_{23}}). \tag{9}$$

   To see (8), note that $\mu_{Xn} = \theta_n \in C_{\mu_{Xn}}(z_{1-\alpha/2})$ is equivalent to $|n^{1/2}(\overline{X}_n - \mu_{Xn})/\widehat{\sigma}_{Xn}| < z_{1-\alpha/2}$ and the probability of this event converges to the probability that $|\xi_{1,h_{23}}| < z_{1-\alpha/2}$ by the CLT and noting that $\sigma_{Xn}/\widehat{\sigma}_{Xn} \to_p 1$. Likewise, $\mu_{Xn} = \theta_n \in C_{\mu_{Y_n}}(z_{1-\alpha/2})$ can be rewritten as $|n^{1/2}(\overline{Y}_n - \mu_{Yn})/\widehat{\sigma}_{Yn} + n^{1/2}(\mu_{Yn} - \mu_{Xn})/\widehat{\sigma}_{Yn}| < z_{1-\alpha/2}$, and the probability of this event converges to $P(|\xi_{2,h_{23}} - h_1/h_{22}| < z_{1-\alpha/2})$. Finally, $\overline{X}_n \geq \overline{Y}_n$ can be rewritten as $\widehat{\sigma}_{Xn} n^{1/2}(\overline{X}_n - \mu_{Xn})/\widehat{\sigma}_{Xn} - n^{1/2}(\mu_{Yn} - \mu_{Xn}) \geq \widehat{\sigma}_{Yn} n^{1/2}(\overline{Y}_n - \mu_{Yn})/\widehat{\sigma}_{Yn}$. Asymptotically, this event occurs when $h_{21}\xi_{1,h_{23}} + h_1 < h_{22}\xi_{2,h_{23}}$. By the joint CLT, the above then implies (8).

Given the probabilities in (8) and (9) only depend on h Assumption LRP is satisfied and the AsyCS is given as the infimum over all h of these probabilities. One can show by simulation that for a given nominal coverage probability $1 - \alpha$ the AsyCS is much smaller.

(iii) Denote by E_n the event $\overline{X}_n + n^{-1/2}\widehat{\sigma}_{Xn} z_{1-\alpha/2} \geq \overline{Y}_n + n^{-1/2}\widehat{\sigma}_{Yn} z_{1-\alpha/2}$ and define $T_n(\theta) = |n^{1/2}(\overline{X}_n - \theta)/\widehat{\sigma}_{Xn}| I(E_n) + |n^{1/2}(\overline{Y}_n - \theta)/\widehat{\sigma}_{Yn}|(1 - I(E_n))$.

Set $E = I(h_{21}\xi_{X,h_{23}} + h_1 + (h_{21} - h_{22}) z_{1-\alpha/2} \geq h_{22}\xi_{Y,h_{23}})$. Under drifting sequences with h_1 finite, we have $T_n(\theta_n) \to_d J_h$, where for $h_1 \geq 0$

$$J_h \sim |\xi_{X,h_{23}}| E + |\xi_{Y,h_{23}} - (h_1/h_{22})|(1 - E) \tag{10}$$

and for $h_1 < 0$

$$J_h \sim |\xi_{X,h_{23}} + (h_1/h_{21})| E + |\xi_{Y,h_{23}}|(1 - E) \tag{11}$$

Note that J_h is continuous. It follows that $\liminf_{n\to\infty} P_{\gamma_{nh}}(T_n(\theta_n) \leq z_{1-\alpha/2}) = J_h(z_{1-\alpha/2})$. AsyCS equals the infimum of $J_h(z_{1-\alpha/2})$ over all h. Simulations indicate that this confidence interval has correct AsyCS.