Midterm Exam, Econ 510
Wednesday 2/26, 10:30-12:00, Walker Building 124
Patrik Guggenberger, Department of Economics, Penn State

Answer all **four** questions for a total of 32 points, 8 points for each question. Closed book exam. No phones or calculators. You can write on front and back sides of each page; if you need more space, there are additional pages at the end of the exam (which you can also use as scratch paper).

1. Suppose $Y_i = X_i^{'}\beta + U_i$ for $i = 1, 2, ..., n$, $\beta \in R^d$, $\{(X_i, U_i) : i \geq 1\}$ are i.i.d. with $E(U_i \mid X_i) = 0$ a.s. and $E(U_i^2 \mid X_i) < \infty$ a.s.

   a) Derive the limiting distribution of the least squares estimator $\widehat{\beta}_{LS}$ for β. [Do not use Theorem 12.1 from the lecture notes here. Derive the limiting distribution from scratch.]

   Solution: (2 point) This question is based on **problem set 2, question 5.** By definition
$$\widehat{\beta}_{LS} = (X'X)^{-1}X'Y,$$
   where $X \in R^{n \times d}$ is the matrix of regressors with i-th row given by $X_i^{'}$ for $i = 1, 2, ..., n$ and analogously for $Y, U \in R^n$. Therefore,
$$n^{1/2}(\widehat{\beta}_{LS} - \beta) = (X'X/n)^{-1}X'U/n^{1/2}.$$
   By a WLLN and the CMT, assuming $E||X_i||^2 < \infty$ and $EX_iX_i' > 0$, we have
$$(X'X/n)^{-1} \to_p (EX_iX_i')^{-1}$$
   and by a CLT for iid random vectors
$$X'U/n^{1/2} \to_d N(0, EU_i^2X_iX_i')$$
   assuming that $E||U_iX_i||^2 < \infty$ and noting that by the LIE $EU_iX_i = E[U_i|X_i]X_i = 0$. Therefore, by Slutsky's theorem
$$n^{1/2}(\widehat{\beta}_{LS} - \beta) \to_d (EX_iX_i')^{-1}N(0, EU_i^2X_iX_i') \sim N(0, \Sigma)$$
   with
$$\Sigma = (EX_iX_i')^{-1}E(U_i^2X_iX_i')(EX_iX_i')^{-1}.$$

   b) Provide an estimator for the variance-covariance matrix of the limiting distribution in a).

   Solution: (1 point) One can use the Eicker–White estimator
$$\widehat{\Sigma}_n = \left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i^{'}\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \widehat{U}_i^2 X_iX_i^{'} \left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i^{'}\right)^{-1},$$
   where $\widehat{U}_i := Y_i - X_i^{'}\widehat{\beta}_{LS}$.

c) Show that the proposed estimator in b) is consistent.

Solution: (2 points) First $\frac{1}{n}\sum_{i=1}^n X_i X_i \to_p E X_i X_i'$ by WLLN. Thus, it is enough to show that $\frac{1}{n}\sum_{i=1}^n \hat{U}_i^2 X_i X_i' \to_p E U_i^2 X_i X_i'$. Note that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \hat{U}_i^2 X_i X_i' - EU_i^2 X_i X_i' &= \frac{1}{n}\sum_{i=1}^n (\hat{U}_i^2 - U_i^2) X_i X_i' + \frac{1}{n}\sum_{i=1}^n U_i^2 X_i X_i - EU_i^2 X_i X_i' \\ &= \frac{1}{n}\sum_{i=1}^n (\hat{U}_i^2 - U_i^2) X_i X_i' + o_p(1).\end{aligned}$$

The last equality comes from WLLN for iid data assuming $E||U_i X_i||^2 < \infty$. Thus, it is enough to show $\frac{1}{n}\sum_{i=1}^n (\hat{U}_i^2 - U_i^2) X_i X_i' \to_p 0$. Note that

$$\begin{aligned}\hat{U}_i - U_i &= Y_i - X_i'\widehat{\beta}_{LS} - Y_i + X_i'\beta_0 = X_i'(\beta_0 - \widehat{\beta}_{LS}), \\ \hat{U}_i &= U_i + X_i'(\beta_0 - \widehat{\beta}_{LS}) = U_i + X_i' O_p(n^{-1/2}), \\ \hat{U}_i^2 - U_i^2 &= 2U_i X_i' O_p(n^{-1/2}) + (X_i' O_p(n^{-1/2}))^2\end{aligned}$$

with the $O_p(n^{-1/2})$ term not depending on i. Therefore

$$\frac{1}{n}\sum_{i=1}^n (\hat{U}_i^2 - U_i^2) X_i X_i' = \frac{1}{n}\sum_{i=1}^n [2U_i X_i' O_p(n^{-1/2}) + (X_i' O_p(n^{-1/2}))^2] X_i X_i' \tag{1}$$

Denote by X_{ij} the j-th component of X_i for $j = 1, ..., d$. By a WLLN $\frac{1}{n}\sum_{i=1}^n U_i X_{ij} X_i X_i' \to_p EU_i X_{ij} X_i X_i' = EX_{ij} X_i X_i' E(U_i|X_i) = 0$ assuming $E||U_i X_{ij} X_i X_i'|| < \infty$ and therefore the first summand in (1) converges to zero in probability. By a WLLN $\frac{1}{n}\sum_{i=1}^n X_{ij} X_{ik} X_i X_i' \to_p EX_{ij} X_{ik} X_i X_i'$ assuming $E||X_{ij} X_{ik} X_i X_i'|| < \infty$ and therefore the second summand in (1) converges to zero in probability.

d) Write down the Wald statistic to test the null hypothesis $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$. Describe the intuition underlying the construction of that statistic.

Solution: (2 points) We are testing $\beta = 0$. Therefore, in the notation of Lecture 14, page 1, $h(\beta) = \beta$ and $h = id$, the identity function and thus $H = I_d$. Therefore, (14.3) becomes

$$\mathcal{W}_n = n\widehat{\beta}_{LS}'(\widehat{B}_n^{-1}\widehat{\Omega}_n\widehat{B}_n^{-1})^{-1}\widehat{\beta}_{LS}.$$

In the particular context here $\widehat{B}_n^{-1}\widehat{\Omega}_n\widehat{B}_n^{-1}$ becomes

$$\widehat{\Sigma}_n = \left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n \widehat{U}_i^2 X_i X_i' \left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)^{-1}.$$

The intuition is given on p.2, first paragraph of p.14. Namely, $\widehat{\beta}_{LS}$ is an estimator of the true β. Therefore, if the true β satisfies the restriction of the null then we would expect $\widehat{\beta}_{LS}$ to do so too and for the given null here, to therefore be "small". Under the null, we therefore expect the Wald statistic to be "small" and we therefore reject if the value of the test statistic is "too big". Too big refers to the appropriate quantile of the limiting null distribution of the test statistic. The weighting matrix in the quadratic form is a consistent estimator of the inverse of the limiting variance covariance matrix. Therefore, given normality, the limiting null distribution is chisquare with d degrees of freedom.

e) Explain how you would construct an (asymptotically valid) 95% confidence region for β based on a Wald test.

Solution: (1 point) Define the confidence region $CR_{95\%}$ as the set of vectors β_0 such that the Wald test at nominal size 5% does not reject the null hypothesis $H_0 : \beta = \beta_0$. This confidence region is asymptotically valid (meaning its asymptotic coverage probability equals 95%) because for the true β_0

$$\begin{aligned} P(\beta_0 &\in CR_{95\%}) = P(\text{Wald does not reject } H_0 : \beta = \beta_0) \\ &= 1 - P(\text{Wald rejects true } H_0 : \beta = \beta_0) \rightarrow 1 - 5\% = 95\% \end{aligned}$$

because the Wald test has correct limiting null rejection probability.

State any additional assumptions you need, if any.

2. Derive the asymptotic distribution of the LM statistic

$$\mathcal{LM}_n = n\frac{\partial}{\partial\theta'}Q_n(\widetilde{\theta}_n)\widetilde{B}_n^{-1}\widetilde{H}_n'(\widetilde{H}_n\widetilde{B}_n^{-1}\widetilde{\Omega}_n\widetilde{B}_n^{-1}\widetilde{H}_n')^{-1}\widetilde{H}_n\widetilde{B}_n^{-1}\frac{\partial}{\partial\theta}Q_n(\widetilde{\theta}_n)$$

under local alternatives $\theta_n = \theta_0 + \lambda/\sqrt{n}$ under suitable high-level assumptions. A sketch is sufficient.

Explain what the quantities $\widetilde{\theta}_n$, $\widetilde{\Omega}_n$, $\widetilde{B}_n$, and $\widetilde{H}_n$ are estimating.

Describe the intuition underlying the construction of the LM statistic.

Solution: The solution is taken from **problem set 3 question 2.**

a) **(6 points)** We assume $\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_n) \to_d N(0, \Omega_0)$, $\tilde{\theta}_n \to_p \theta_0$, and $\tilde{B}_n \to_p B_0$ and $\tilde{\Omega}_n \to_n \Omega_0$ to hold under the local alternatives. We will prove that $\mathcal{LM}_n$ converges in distribution to a $\chi^2(\delta)$, where

$$\delta = \lambda' H'(HB_0^{-1}\Omega_0 B_0 H')^{-1}H\lambda$$

under the local alternatives.

By a mean value expansion centered at θ_n

$$\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\tilde{\theta}_n) = \sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_n) + \sqrt{n}\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta_n^*)(\tilde{\theta}_n - \theta_n) \text{ and}$$

$$0 = h(\tilde{\theta}_n) = h(\theta_n) + \frac{\partial}{\partial\theta}h(\theta_n^+)(\tilde{\theta}_n - \theta_n).$$

Note that by $\theta_n = \theta_0 + \lambda/\sqrt{n}$ it follows that

$$\sqrt{n}(h(\theta_n) = \sqrt{n}(h(\theta_n) - h(\theta_0)) = \frac{\partial}{\partial\theta}h(\theta_n^{**})\sqrt{n}(\tilde{\theta}_n - \theta_n) = \frac{\partial}{\partial\theta}h(\theta_n^{**})\lambda \to_p H\lambda.$$

Then

$$\begin{aligned}
& \frac{\partial}{\partial\theta}h(\theta_n^+)[\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta_n^*)]^{-1}\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\tilde{\theta}_n) \\
= & \frac{\partial}{\partial\theta}h(\theta_n^+)[\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta_n^*)]^{-1}[\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_n) + \sqrt{n}\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta_n^*)(\tilde{\theta}_n - \theta_n)] \\
= & \frac{\partial}{\partial\theta}h(\theta_n^+)[\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta_n^*)]^{-1}\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_n) - \sqrt{n}h(\theta_n) \\
\to & {}_d Z_0 + H\lambda,
\end{aligned}$$

where $Z_0 \sim N(0, HB_0^{-1}\Omega_0 B_0 H')$. One can show $\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\tilde{\theta}_n) = O_p(1)$. The above then also implies that

$$\widetilde{H}_n\widetilde{B}_n^{-1}\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\tilde{\theta}_n) \to_d Z_0 + H\lambda.$$

Then

$$[\widetilde{H}_n\widetilde{B}_n^{-1}\widetilde{\Omega}_n\widetilde{B}_n^{-1}\widetilde{H}_n']^{-1/2}\widetilde{H}_n\widetilde{B}_n^{-1}\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\tilde{\theta}_n) \to_d [HB_0^{-1}\Omega_0 B_0^{-1}H']^{-1/2}(Z_0 + H\lambda),$$

where the RHS is distributed normally with identity covariance matrix and mean vector $[HB_0^{-1}\Omega_0 B_0^{-1}H']^{-1/2}H\lambda$. Therefore, the LM statistic, being the inner product of the LHS converges to the noncentral chi squared given above.

b) (1 point) $\widetilde{\theta}_n$ represents the restricted EE for θ_0, $\widetilde{\Omega}_n$ represents a consistent estimator under the null for Ω using $\widetilde{\theta}_n$ that appears in CF(iii), the limiting variance covariance matrix of $\partial Q_n(\theta_0)/\partial\theta$, $\widetilde{B}_n$ represents a consistent estimator under the null for B_0 that appears in CF(iv), where B_0 is the probability limit of the Hessian $\partial^2 Q_n(\theta_0)/\partial\theta'\partial\theta$, and $\widetilde{H}_n$ is a consistent estimator under the null for $H(\theta_0)$ where $H(\theta) = \partial h(\theta)/\partial\theta$ and h is the function that appears in the formulation of the null hypothesis.

c) (1 point) By definition of the EE and the FOC we know that $\frac{\partial}{\partial\theta}Q_n(\widehat{\theta}_n)$ is approximately zero. If the null is true the restricted $\widetilde{\theta}_n$ and unrestricted estimator $\widehat{\theta}_n$ should be close to each other and therefore, under the null $\frac{\partial}{\partial\theta}Q_n(\tilde{\theta}_n)$ should be close to zero. The LM statistic is a a quadratic form in the renormalized $\frac{\partial}{\partial\theta}Q_n(\tilde{\theta}_n)$ that is $O_p(1)$ under the null, but goes off to infinity under the (fixed) alternative.

3. Assume $W_i \sim$ i.i.d. $N(\theta_1, \theta_2)$ for $i = 1, ..., n$ for some unknown values $\theta_1 \in R$, $\theta_2 > 0$. The objective is to test the null hypothesis $H_0 : \theta_1 = 0$ versus $H_1 : \theta_1 \neq 0$.

   a) Explain in detail how to implement the test using the parametric bootstrap.

   b) Define what consistency of the bootstrap means and prove that the parametric bootstrap considered in a) is consistent.

   c) Assume $H_1 : \theta_1 \neq 0$ is true. Discuss what happens to the power of the test as $n \to \infty$. Provide an argument for your claim.

   d) Derive the power of the test under local alternatives.

   Solutions. Each part carries two points. **a)** To test the null H_0 we need to pick a nominal size $\alpha > 0$, define a test statistic, and a critical value. The test statistic can be taken as a t-statistic

$$t_n = |n^{1/2} \frac{\widehat{\theta}_1}{sd(\widehat{\theta}_1)}|,$$

   where $\widehat{\theta}_1 = \overline{W}_n = n^{-1} \sum_{i=1}^n W_i$ and, given that under the null

$$n^{1/2} \widehat{\theta}_1 \to_d N(0, \theta_2),$$

   we can take $sd(\widehat{\theta}_1) = \widehat{\theta}_2 = n^{-1} \sum_{i=1}^n (W_i - \overline{W}_n)^2$.

   The bootstrap critical value is obtained as the $1 - \alpha$ quantile of t_n^* where

$$t_n^* = |n^{1/2} \frac{\widehat{\theta}_1^* - \widehat{\theta}_1}{sd(\widehat{\theta}_1^*)}|$$

   denotes the recentered t-statistic and the estimators are defined as before but evaluated on resampled data. In the case of parametric bootstrap, the B resampled data sets W_{ib}^* for $i = 1, ..., n$ and $b = 1, ..., B$ are drawn iid from $N(\widehat{\theta}_1, \widehat{\theta}_2)$. Here, $sd(\widehat{\theta}_1^*) = n^{-1} \sum_{i=1}^n (W_{ib}^* - \overline{W}_n)^2$.

   b) We defined consistency in the lecture as the bootstrap test statistic having the same limiting distribution conditional on the data (wp1) as the original test statistic under the null hypothesis. By part a) and the WLLN, under the null, $t_n^* \to_d |N(0, 1)|$. We therefore have to show that $t_n^* \to_d |N(0, 1)|$ under the null, conditional on the data (wp1). The proof is virtually identical to what we did in class for the nonparametric bootstrap.

   Note that trivially $E^* W_{ib}^* = \widehat{\theta}_1$ and $E^*(W_{ib}^* - \widehat{\theta}_1)^2 = \widehat{\theta}_2$ given the bootstrap data is generated from $N(\widehat{\theta}_1, \widehat{\theta}_2)$. Define the sample average $\overline{W}_b^* = n^{-1} \sum_{i=1}^n W_{ib}^*$ (which then also satisfies $E^* \overline{W}_b^* = \widehat{\theta}_1$).

   Note that, therefore, $W_{ib}^* - \widehat{\theta}_1 = W_{ib}^* - E^* W_{ib}^*$ are independent zero mean variables with variance equal to $\widehat{\theta}_2 = n^{-1} \sum_{i=1}^n (W_i - \overline{W}_n)^2$ which by the

SLLN converges to θ_2 a.s.. Let's condition on a sequence of data for which the convergence holds. Then, by a CLT and the WLLN

$$t_n^* = |n^{1/2}\frac{\widehat{\theta}_1^* - \widehat{\theta}_1}{sd(\widehat{\theta}_1^*)}| = |\frac{n^{-1/2}\sum_{i=1}^n (W_{ib}^* - E^* W_{ib}^*)}{n^{-1}\sum_{i=1}^n (W_{ib}^* - E^* W_{ib}^*)^2}| \to_d |N(0,1)|$$

as desired.

c) The power of the test will converge to 1. First consider what happens to the test statistic

$$t_n = |n^{1/2}\frac{\widehat{\theta}_1 - \theta_1}{sd(\widehat{\theta}_1)} + n^{1/2}\frac{\theta_1}{sd(\widehat{\theta}_1)}|.$$

The first summand $n^{1/2}\frac{\widehat{\theta}_1 - \theta_1}{sd(\widehat{\theta}_1)}$ converges in distribution to a normal while the second summand diverges to (plus or minus) infinity. Therefore, $t_n \to \infty$. However, the bootstrap statistic remains $O_p(1)$ by the same proof as in b).

d) If $\theta_1 = n^{-1/2}h$ for some nonzero $h \in R$, then, from c)

$$t_n = |n^{1/2}\frac{\widehat{\theta}_1 - \theta_1}{sd(\widehat{\theta}_1)} + n^{1/2}\frac{\theta_1}{sd(\widehat{\theta}_1)}| \to |N(\frac{h}{\theta_2}, 1)|$$

while the limiting distribution of the bootstrap statistic is again as under b) and c). Therefore, power converges to $P(Z_h > z_{1-\alpha/2})$, where $Z_h \sim |N(\frac{h}{\theta_2}, 1)|$.

4. 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

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

a) Provide estimators for V_0 and Γ and discuss their consistency under appropriate conditions.

b) Consider the example of quantile Regression, $(Y_i, X_i')'$ i.i.d.

$$Y_i = X_i'\theta_0 + U_i,$$

where the τ-quantile of U_i conditional on X_i equals 0 for some $\tau \in (0,1)$.

Define the quantile estimator of θ_0. State high level assumptions for its consistency, **EE3**(ii). Provide primitive conditions for the high level assumptions in the context of the quantile regression example.

Solutions: Each part carries four points. **a)** This is taken straight from the Lecture Notes. See (16.24) and the discussion below for the estimator for V_0 and its consistency and (16.25)-(16.26) for the estimator for Γ and its consistency.

b) The quantile estimator is defined as the minimizer of the stochastic criterion function defined in (16.28) (up to an $o_p(1)$ term)[1]

$$Q_n(\theta) = n^{-1} \textstyle\sum_{i=1}^n \rho_\tau(Y_i - X_i'\theta)$$

for ρ_τ being the check function.

From Theorem 11.1 we know that the estimator is consistent if ID and U-$WCON$ hold. Define $Q(\theta)$ as $E\rho_\tau(Y_i - X_i'\theta)$, that is the pointwise probability limit of $Q_n(\theta)$ (by the WLLN assuming $E\rho_\tau(Y_i - X_i'\theta) < \infty$ holds true).

Sufficient conditions for ID are given by $ID1$ namely a compact parameter space Θ, $Q(\theta)$ continous and θ_0 uniquely minimizing $Q(\theta)$. We also know by Theorem 11.3 that U-$WCON$ holds if the data are iid, $\rho_\tau(Y_i - X_i'\theta)$ is continuous in θ (which holds here), $E\sup_{\theta\in\Theta} \rho_\tau(Y_i - X_i'\theta) < \infty$, and the parameter space is compact. [Up to here was sufficient for full score]

Regarding θ_0 minimizing $Q(\theta)$, taking FOC, and interchaning the order of integration and differentiation (note that $\rho_\tau(Y_i - X_i'\theta)$ is differentiable unless $Y_i - X_i'\theta = 0$ which occurs with probability 0, see the discussion below (16.3)) we obtain $g(\theta) = 0$ for the function g defined in (16.30). In (16.31) it has been established that indeed $g(\theta_0) = 0$.

[1] To work out the limiting distribution of the estimator we took the first order condition, see (16.29), and then went on interpreting the quantile estimator as the solution to a GMM framework.

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) $\widehat{\theta}_n \rightarrow_p \theta_0$.