Final Exam, Econ 510 **YOUR NAME:**

Wednesday 5/6/2026, 10:10am-12:00pm, 107 Chambers

Patrik Guggenberger, Department of Economics, Penn State

Answer all **five** questions for a total of 50 points. Closed book exam, no calculators or cell phones allowed. You can write on the 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). GOOD LUCK!

1. **(13 points)** The model is $y_i = x_i'\beta + \varepsilon_i$, $i = 1, ..., n$, with iid data, where x_i is potentially endogenous with reduced form $x_i' = Z_i'\Pi + v_i'$ where Z_i is a vector of IVs. Define

$$\begin{aligned}
g_i(\beta) \quad &: \quad = (y_i - x_i'\beta)Z_i, \ G_i := (\partial g_i/\partial\beta)(\beta) = -Z_i x_i', \\
\widehat{g}(\beta) \quad &: \quad = \sum_{i=1}^{n} g_i(\beta)/n, \ \widehat{\Omega}(\beta) := \sum_{i=1}^{n} g_i(\beta)g_i(\beta)'/n, \\
D(\beta) \quad &: \quad = \sum_{i=1}^{n} (\widehat{g}(\beta)'\widehat{\Omega}(\beta)^{-1}g_i(\beta) - 1)G_i/n, \\
LM_{CUE}(\beta) \quad &: \quad = n\widehat{g}(\beta)'\widehat{\Omega}(\beta)^{-1}D(\beta)[D(\beta)'\widehat{\Omega}(\beta)^{-1}D(\beta)]^{-1}D(\beta)'\widehat{\Omega}(\beta)^{-1}\widehat{g}(\beta).
\end{aligned}$$

Show that for $n \to \infty$ we have $LM_{CUE}(\beta_0) \to_d \chi^2_{\dim \beta_0}$ under weak IV asymptotics $\Pi = \Pi_n = n^{-1/2}C$ for fixed C. To do so,

a) **(6 points)** work out the joint limiting normal distribution, $(\overline{g}', vec(\overline{D})')'$ say, of

$$n^{1/2}(\widehat{g}(\beta_0)', vec(D(\beta_0))')'$$

and show that asymptotically $n^{1/2}\widehat{g}(\beta_0)$ and $n^{1/2}vec(D(\beta_0))$ are independent.

b) **(4 points)** Show that the limiting distribution of $LM_{CUE}(\beta_0)$ conditional on $vec(\overline{D})$ is $\chi^2_{\dim \beta_0}$. Conclude that the same result therefore holds unconditionally.

State any additional assumptions you use.

c) **(1 point)** Consider testing the null hypothesis $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ at nominal size $\alpha > 0$. Consider the test that rejects when $LM_{CUE}(\beta_0) > \chi^2_{\dim \beta_0, 1-\alpha}$, where the latter denotes the $(1 - \alpha)$-quantile of $\chi^2_{\dim \beta_0}$. What is the advantage (if any) of that test over a textbook Wald test?

d) **(2 points - split up as .5 points for naming the CLR test and 1.5 points for the explanation)** Assume $\dim \beta_0 = 1$ and assume conditional homoskedasticity. Which test would you recommend under these assumptions and why?

Solutions. Parts a-b) were taken from problem set 6 question 3.

a) Rearranging terms and normalizing each part properly we obtain

$$\begin{aligned} n^{1/2}D(\beta_0) \quad &: \quad = n^{1/2}\sum_{i=1}^{n}(\widehat{g}(\beta_0)'\widehat{\Omega}(\beta_0)^{-1}g_i(\beta_0)-1)G_i/n \\ &= \quad -n^{-1/2}\sum_{i=1}^{n}G_i + n^{-1}\sum_{i=1}^{n}G_i g_i(\beta_0)'\widehat{\Omega}(\beta_0)^{-1}n^{1/2}\widehat{g}(\beta_0) \\ &= \quad \frac{1}{n^{1/2}}\sum_{j=1}^{n}Z_j x_j' - n^{-1}\sum_{i=1}^{n}(Z_i x_i')Z_i'\varepsilon_i\widehat{\Omega}(\beta_0)^{-1}\frac{1}{n^{1/2}}\sum_{j=1}^{n}Z_j\varepsilon_j, \end{aligned}$$

where in the last line we use $G_i = -Z_i x_i'$ and $n^{1/2}\widehat{g}(\beta_0) = \frac{1}{n^{1/2}}\sum_{j=1}^{n}Z_j\varepsilon_j$.

Note that

$$Z_i x_i' = Z_i Z_i'\Pi + Z_i v_i'.$$

Therefore, using $\Pi = n^{-1/2}C$, $vec(n^{1/2}D(\beta_0))$ consists of the p column k-vectors, D_l^* say,

$$n^{-1}\sum_{j=1}^{n}Z_j Z_j' C_l + \frac{1}{n^{1/2}}\sum_{j=1}^{n}Z_i v_{il} - n^{-1}\sum_{i=1}^{n}Z_i v_{il}Z_i'\varepsilon_i\widehat{\Omega}(\beta_0)^{-1}\frac{1}{n^{1/2}}\sum_{j=1}^{n}Z_j\varepsilon_j + o_p(1)$$

for $l = 1, ..., p$, where C_l denotes the l-th column of C and v_{il} denotes the l-th component of v_i. The $o_p(1)$ term comes from the term

$$n^{-3/2}\sum_{i=1}^{n}(Z_i Z_i' C_l)Z_i'\varepsilon_i\widehat{\Omega}(\beta_0)^{-1}\frac{1}{n^{1/2}}\sum_{j=1}^{n}Z_j\varepsilon_j$$

under the assumption that $E|Z_{ip}Z_{il}Z_{im}\varepsilon_i| < \infty$ for $p, l, m = 1, ..., k$, $E||Z_j\varepsilon_j||^2 < \infty$, and $\Omega > 0$, by the WLLN and the CLT (Note that this term is even $O_p(n^{-1/2})$). By the WLLN we have

$$n^{-1}\sum_{i=1}^{n}Z_i v_{il}Z_i'\varepsilon_i\widehat{\Omega}(\beta_0)^{-1} \to_p EZ_i Z_i'\varepsilon_i v_{il}\Omega^{-1} =: \Lambda,\ n^{-1}\sum_{j=1}^{n}Z_j Z_j' C_l \to_p EZ_j Z_j' C_l.$$

By the CLT

$$(\frac{1}{n^{1/2}}\sum_{j=1}^{n}Z_j'\varepsilon_j, \frac{1}{n^{1/2}}\sum_{j=1}^{n}Z_i' v_{il})' \to_d \zeta \sim N(0, \Delta),$$

where $\Delta \in R^{2k\times 2k}$ with lower right block equal to $Ev_{il}^2 Z_i Z_i'$, upper left block equal to $E\varepsilon_i^2 Z_i Z_i'$ and off diagonal blocks equal to $E\varepsilon_i v_{il}Z_i Z_i'$.

Thus, because

$$(n^{1/2}\widehat{g}(\beta_0), D_l^*) = \begin{pmatrix} I & 0 \\ -\Lambda & I \end{pmatrix}(\frac{1}{n^{1/2}}\sum_{j=1}^{n}Z_j'\varepsilon_j, \frac{1}{n^{1/2}}\sum_{j=1}^{n}Z_i' v_{il})' + o_p(1)$$

we obtain

$$(n^{1/2}\widehat{g}(\beta_0), D_l^*) \to_d N((0', (EZ_j Z_j' C_l)')', \Psi) \qquad (1)$$

with $\Psi \in R^{2k\times 2k}$ with lower right block equal to

$$Ev_{il}^2 Z_i Z_i' - E\varepsilon_i v_{il} Z_i Z_i' \Omega^{-1} E\varepsilon_i v_{il} Z_i Z_i',$$

upper left block equal to $E\varepsilon_i^2 Z_i Z_i'$ and off diagonal blocks equal to

$$-E\varepsilon_i^2 Z_i Z_i' \Omega^{-1} E\varepsilon_i v_{il} Z_i Z_i' - E\varepsilon_i v_{il} Z_i Z_i' = 0$$

(because $\Omega = E\varepsilon_i^2 Z_i Z_i'$). The latter implies that $n^{1/2}\widehat{g}(\beta_0)$ and $n^{1/2}vec(D(\beta_0))$ are asymptotically independent because the convergence in (1) holds jointly for all $l = 1, ..., p$.

b) Recall

$$\begin{aligned} & LM_{CUE}(\beta) \\ = \ & n^{1/2}\widehat{g}(\beta)'\widehat{\Omega}(\beta)^{-1} n^{1/2} D(\beta)[n^{1/2} D(\beta)'\widehat{\Omega}(\beta)^{-1} n^{1/2} D(\beta)]^{-1} n^{1/2} D(\beta)'\widehat{\Omega}(\beta)^{-1} n^{1/2}\widehat{g}(\beta) \end{aligned}$$

and thus, by a)

$$LM_{CUE}(\beta) \rightarrow_d \overline{LM} \sim \xi'\widehat{\Omega}(\beta)^{-1/2}\overline{D}[\overline{D}'\widehat{\Omega}(\beta)^{-1}\overline{D}]^{-1}\overline{D}'\widehat{\Omega}(\beta)^{-1/2}\xi,$$

where $\xi \sim \Omega^{-1/2}\overline{g}$ is multivariate normal with zero mean and identity variance covariance matrix and independent of $\overline{D}$ whose stacked columns are also normally distributed. It is easily shown that with probability one $\overline{D}$ has full column rank. Conditional on $\overline{D}$, $\xi \sim N(0, I)$ and by the CMT $\overline{LM} \sim \chi^2_{\dim \beta_0}$ (the proof of that step is a simplified version of the proof of question 4i) of problem set 9). But if $\overline{LM} \sim \chi^2_{\dim \beta_0}$ conditional on $\overline{D}$ then also unconditionally.

c) We have noted in class that the Wald test has asymptotic size equal to 1 for any nonzero nominal size for a paramter space that does not put any restrictions on the weakness of the IVs. In contrast, by part b), the $LM_{CUE}(\beta)$ test has asymptotic size equal to the nominal size given that under drifting weak IV sequences (and also under strong IV) it has limiting null rejection probability equal to the nominal size.

d) Under these assumptions one should use the conditional likelihood ratio test by Moreira (Ecta, 2003). We have discussed in class the result in Andrews, Moreira, and Stock (Ecta, 2006) that states that (under normality and fixed IVs) the power curve of the CLR test gets uniformly close to (a certain) two-sided power envelope in the class of similar invariant tests. This result has immediate implications on the power properties of the CLR test under weak IV asymptotics.

2. (7 points) Let X_i ~iid with unknown mean $\beta \in R$ and variance 1, $i = 1, ..., n$. Assume we are interested in testing

$$H_0 : \beta = 0 \text{ vs } H_1 : \beta > 0.$$

based on the statistic $n^{1/2}\overline{X}_n$, where $\overline{X}_n = \sum_{i=1}^n X_i/n$. Regarding a critical value, two resampling schemes are suggested.

1) Let $X_{ib}^* \sim \text{iid}\widehat{F}_n$, for $i = 1, ..., n$ and $b = 1, ..., B$ (for some user chosen B, say 1000) where $\widehat{F}_n$ denotes the EDF of the data and define $\overline{X}_{nb}^* = \sum_{i=1}^n X_{ib}^*/n$. Take as the critical value the $1 - \alpha$-quantile of $n^{1/2}\overline{X}_{nb}^*$, $b = 1, ..., B$.

2) Choose $s = n^{1/2}$. Consider all the different subsets S_j, $j = 1, ..., \binom{n}{s}$, of $\{X_1, ..., X_n\}$ with s elements and for each subset S_j calculate the sample average $\overline{X}_{sj} = \sum_{i \in S_j} X_i/s$. Take as the critical value the $1 - \alpha$-quantile of $s^{1/2}\overline{X}_{sj}$, $j = 1, ..., \binom{n}{s}$.

What can you say about the asymptotic null rejection probability of the two procedures? Consistency? Would you recommend any of the two procedures? Explain.

(3.5 points for the discussion of 1) and 3.5 points for the discussion of 2))

Solutions: From the lecture, we know that method 2), called subsampling, controls the asymptotic null rejection probability and is also consistent given that the blocksize has been chosen as $n^{1/2}$ which goes off to infinity but slower so then the sample size n in the sense that $n^{1/2}/n \to 0$. Compared to simply using the $1-\alpha$-quantile of a standard normal as the critical value, subampling is computationally more involved and has worse power. Namely, note that under a fixed alternative $\beta \neq 0$ the subsampling critical value goes off to infinity at rate $s^{1/2}$ whereas the $1-\alpha$-quantile of a standard normal critical value (trivially) remains constant.

However, method 1) is problematic. Note that the nonparametric bootstrap would generate a critical value from the RECENTERED statistics $\sum_{i=1}^{n}(X_{ib}^{*} - \overline{X}_n)/n^{1/2}$ and not from $\sum_{i=1}^{n} X_{ib}^{*}/n^{1/2}$ as suggested here. We have shown in class that $\sum_{i=1}^{n}(X_{ib}^{*} - \overline{X}_n)/n^{1/2}$ has the same limiting distribution as $n^{1/2}\overline{X}_n$ under the null (namely a N(0,1)) conditional on the data. That implies that

$$n^{-1/2}\sum_{i=1}^{n} X_{ib}^{*} = n^{-1/2}\sum_{i=1}^{n}(X_{ib}^{*} - \overline{X}_n) + n^{1/2}\overline{X}_n$$

conditional on the data, asymptotically, follows a normal distribution with variance 1 whose mean equals the limit of $n^{1/2}\overline{X}_n$. That is, in particular, asymptotically, with 50% probability the mean will be negative and the resulting critical value will be too small relative to the $1-\alpha$-quantile of a standard normal critical value leading to overrejection of the null hypothesis.

Given the above analysis, the **first method should not be used at all** and the **second one is dominated by the delta method**. Note that a correctly implemented nonparametric bootstrap method offers higher order improvements..

3. (10 points) In the Heckit model (Heckman, 1979)

$$\begin{aligned} y^* &= x'\theta + \varepsilon, \\ y &= dy^*, \\ d &= 1(x'\pi_1 + z'\pi_2 \geq -\eta), \end{aligned}$$

assume $(\varepsilon, \eta) \perp (x, z)$ and $(\varepsilon, \eta)'\tilde{} N(0_2, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon\eta} & 1 \end{pmatrix})$. The objective is to identify π_1, π_2, θ, and $\sigma_{\varepsilon\eta}$.

a) **(1 point)** Provide some discussion about the relevance of that model in applications in Economics.

b) **(1 point)** Discuss the normalization to 1 of the variance of η.

c) **(3 points)** Show that

$$E(y|x, z, d = 1) = x'\theta + \sigma_{\varepsilon\eta}\lambda(x'\pi_1 + z'\pi_2),$$

where $\lambda(s) = \phi(s)/\Phi(s)$.

d) **(3 points)** Identify π_1 and π_2 from a probit regression of d onto x and z.

e) **(2 points)** Specify conditions under which one can identify θ and $\sigma_{\varepsilon\eta}$. State any additional restriction you impose.

Solutions. See problem set 9 question 1 for parts c)-e). Regarding a) y^* **is in general a latent variable and is only observed when** d **equals 1, that is if** $x'\pi_1 + z'\pi_2 \geq -\eta$. **For example, wages are only observed if a** worker is employed. Regarding b) note that because $x'\pi_1 + z'\pi_2 \geq -\eta$ is invariant to multiplication by a positive constant, π_1 and π_2 can only be identified through some sort of normalization. The normalization employed here is to restrict the variance of η to 1..

4. (11 points) Let $\alpha \in (0,1)$. Let $X_i \sim$ iid with mean μ and variance 1, $i = 1, ..., n$, where $\mu \in R^+ = \{y \in R, y \geq 0\}$. The MLE of μ is $\widehat{\mu}_n = \max\{\overline{X}_n, 0\}$, where $\overline{X}_n = \sum_{i=1}^n X_i/n$. Define a confidence interval for μ of nominal coverage probability $1-\alpha$ as

$$CI_{1-\alpha,n} = \{\mu \geq 0 \text{ st } n^{1/2}(\widehat{\mu}_n - \mu) \leq z_{1-\alpha}\}, \tag{2}$$

where $z_{1-\alpha}$ denotes the $1-\alpha$ quantile of a standard normal distribution.

a) **(1 points)** Show that $n^{1/2}(\widehat{\mu}_n - \mu) \rightarrow_d Z$ if $\mu > 0$ as $n \rightarrow \infty$ where $Z \sim N(0,1)$ and $n^{1/2}(\widehat{\mu}_n - \mu) \rightarrow_d \max\{Z, 0\}$ if $\mu = 0$.

b) **(1 points)** Consider a drifting sequence $\mu = \mu_n$ for the true mean such that $n^{1/2}\mu_n \rightarrow h$ for some $h \in [0, \infty]$. Under such a sequence, derive the limiting distribution of $n^{1/2}(\widehat{\mu}_n - \mu_n)$.

c) **(1 points)** Derive a formula for the limiting coverage probability of $CI_{1-\alpha,n}$ under such sequences.

d) **(3 points)** Derive a simple expression for the asymptotic confidence size of $CI_{1-\alpha,n}$.

e) **(1 points)** In case the procedure has asymptotic confidence size smaller than $1-\alpha$ can you suggest a modified version that has correct asymptotic confidence size?

f) **(2 points)** Explain how one would use (a textbook version) of the nonparametric bootstrap to produce critical values in (2) (rather than using $z_{1-\alpha}$). Can you say anything about the asymptotic confidence size of the resulting confidence interval?

g) **(2 points)** Redo part f) using subsampling rather than the nonparametric bootstrap.

Solutions

This question is closely related to problem set 8 questions 2 and 4.

a) $n^{1/2}(\widehat{\mu}_n - \mu) \to_d Z$ if $\mu > 0$ follows from the CLT. When $\mu = 0$ note that $n^{1/2}(\widehat{\mu}_n - \mu) = \max\{n^{1/2}\overline{X}_n, 0\} \to_d \max\{Z, 0\}$ by the CLT and the CMT.

b) $n^{1/2}(\widehat{\mu}_n - \mu) = \max\{n^{1/2}\overline{X}_n, 0\} - n^{1/2}\mu = \max\{n^{1/2}(\overline{X}_n - \mu), -n^{1/2}\mu\} \to_d \max\{Z, -h\}$ by the CLT and the CMT.

c) $P_{\mu_n}(\mu_n \in CI_{1-\alpha,n}) = P_{\mu_n}(n^{1/2}(\widehat{\mu}_n - \mu_n) \leq z_{1-\alpha}) \to P(\max\{Z, -h\} \leq z_{1-\alpha})$.

d) From d) and the theory we developed in class (asymptotic confidence since is the worst limiting coverage probability under sequences of the type considered in b))

$$AsyCs = \inf_{h \in [0,\infty]} P(\max\{Z, -h\} \leq z_{1-\alpha}).$$

Note that $\max\{Z, -h\}$ is decreasing as h increases and when $h = \infty$ we have $P(\max\{Z, -h\} \leq z_{1-\alpha}) = P(Z \leq z_{1-\alpha}) = 1 - \alpha$. It follows that

$$AsyCs = P(\max\{Z, 0\} \leq z_{1-\alpha}).$$

Note that if $z_{1-\alpha} \geq 0$ (that is if $\alpha \leq .5$) it follows that $P(\max\{Z, 0\} \leq z_{1-\alpha}) = P(Z \leq z_{1-\alpha}) = 1 - \alpha$. That is, for nominal coverage probabilities typically employed in practice, $AsyCs$ equals the nominal coverage probability.

If $z_{1-\alpha} < 0$ (that is if $\alpha > .5$ - which is not commonly used in applied work) it follows that $P(\max\{Z, 0\} \leq z_{1-\alpha}) = 0 < 1 - \alpha$. Therefore, in that case, the worst coverage probability is 0%!

e) The modification is only needed when $\alpha > .5$. For any critical value c, by the above, the asymptotic confidence size equals $P(\max\{Z, 0\} \leq c)$. In order to get coverage probability equal to $1 - \alpha$ one needs to take c as the $1 - \alpha$-quantile of the distribution $\max\{Z, 0\}$.

f) On problem set 8 question 2 it was shown that when $\mu = 0$ for some $0 < c < \infty$ the bootstrapped statistic is bounded above by a statistic that along a subsequence n_k (conditional on the data) converges in distribution to $\max\{Z - c, 0\}$. That is, along that subsequence the bootstrap critical value is nonbigger than the $(1 - \alpha)$-quantile of $\max\{Z - c, 0\}$ when in fact the $(1 - \alpha)$-quantile of $\max\{Z, 0\}$ is needed to control the limiting coverage probability. Given that the $(1 - \alpha)$-quantile of $\max\{Z - c, 0\}$ is strictly smaller than the $(1 - \alpha)$-quantile of $\max\{Z, 0\}$ for any $\alpha < 1/2$ the bootstrap CI has asymptotic confidence size smaller than the nominal coverage probability in that case. (When $\alpha \leq 1/2$ the $(1 - \alpha)$-quantile of both $\max\{Z - c, 0\}$ and of $\max\{Z, 0\}$ are both equal to 0.)

g) As in problem set 8 questions 4, for sequences

i) $n^{1/2}\mu_n \to h$ for some $h \in [0,\infty)$, the subsampling critical value will converge to the $1-\alpha$-quantile of the distribution $\max\{Z,0\}$ whereas the test statistic converges to $\max\{Z,-h\}$.

ii) $n^{1/2}\mu_n \to \infty$ and $b^{1/2}\mu_n \to g$ for some $g \in [0,\infty]$, the subsampling critical value will converge to the $1-\alpha$-quantile of the distribution $\max\{Z,-g\}$ whereas the test statistic converges to Z.

Therefore, denoting the $1-\alpha$-quantile of the distribution $\max\{Z,-g\}$ by $c_{g,1-\alpha}$,

$$AsyCs = \min\{\inf_{h\in[0,\infty]} P(\max\{Z,-h\} \le c_{0,1-\alpha}), \inf_{g\in[0,\infty]} P(Z \le c_{g,1-\alpha})\}.$$

As above, $\inf_{h\in[0,\infty]} P(\max\{Z,-h\} \le c_{0,1-\alpha}) = P(\max\{Z,0\} \le c_{0,1-\alpha}) = 1-\alpha$. Also, $\inf_{g\in[0,\infty]} P(Z \le c_{g,1-\alpha}) = P(Z \le c_{\infty,1-\alpha}) = P(Z \le z_{1-\alpha}) = 1-\alpha$. Therefore, $AsyCs = 1-\alpha$..

5. (9 points) Consider the relation $y = X\theta^* + w \in R^n$ with $X \in R^{n \times p}$ where the objetive is estimation of θ^* after observing the data (y, X). Possibly, p is much larger than n.

a) **(2 points)** Define the LASSO estimator. What do we mean by the "oracle property" of LASSO?

b) **(3 points)** Present a set of high level assumptions for given θ^*, the penalty term λ_n used in the implementation of LASSO, X, and w such that the oracle property is satisfied. (A formal presentation is best but an intuitive verbal description is sufficient). Critically discuss these assumptions.

c) **(2 points)** You did finite sample simulations for the LASSO estimator. Describe the performance of LASSO relative to the Ridge estimator under various choices of the DGP.

d) **(2 points)** Show that the Ridge estimator can be calculated via simple OLS regression when augmenting y and X appropriately. Namely, replace y by $(y', 0_p')'$ and X by $\binom{X}{\sqrt{\lambda_n} I_p}$.

Solutions:

a) See Wainwright's textbook (7.18) (or class notes). For the oracle property, see Theorem 7.21 in the textbook - **no false inclusion** and **no false exlusion**.

b) See the classnotes: we require a **lower bound on regularization**, we require the **smallest eigenvalue of the sample covariance submatrix indexed by** S to be bounded away from zero, and **mutual incoherence** (for the latter two conditions, see Wainwright, Assumptions (A3)-(A4), p. 219), and finally, we require the nonzero slope coefficients to be sufficiently bounded away from zero.

Critical discussion: regularization poses a delicate tradeoff - it reduces the variability of the estimator but increases the bias. Getting that tradeoff right in practice is nontrivial (cross-validation is one option).

As discussed in class it is clear that the smallest eigenvalue condition is a necessary condition for identification of the nonzero slope coefficients. The mutual incoherence condition imposes an "approximate orthogonality" between the covariate vectors corresponding to nonzero slope coefficients and covariate vectors correponsding to zero slope coefficients.

The most crucial condition is the requirement that the nonzero slope coefficients are "well separated" from zero. This is an ad-hoc assumption that may or may not be true in practice

c) This refers to problem set 7, question 4. LASSO, leads to a large percentage of correct zero estimates (as predicted by the Oracle property) if the assumptions discussed above are satisfied - Ridge on the other hand does not share the Oracle property. However, if the slope coefficients are not separated far enough from zero then LASSO may wrongly assign them the value 0. With regards to the loss functions (average prediction errors in various norms) considered, Ridge performs better than LASSO when the nonzero slope coefficients are small (that is when one of the key assumptions for the Oracle property fails), otherwise LASSO typically outperformed Ridge.

d)**:** The OLS estimator in a linear regression where y and X are replaced by $\overline{y} := (y', 0_p')'$ and $\overline{X} := \binom{X}{\sqrt{\lambda_n} I_p}$, respectively, reads

$$\begin{aligned}
\widehat{\theta} &= (\overline{X}'\overline{X})^{-1}\overline{X}'\overline{y} \\
&= (\begin{pmatrix} X \\ \sqrt{\lambda_n} I_p \end{pmatrix}' \begin{pmatrix} X \\ \sqrt{\lambda_n} I_p \end{pmatrix})^{-1} \begin{pmatrix} X \\ \sqrt{\lambda_n} I_p \end{pmatrix}' (y', 0_p')' \\
&= (X'X + \lambda_n I_p)^{-1} X'y
\end{aligned}$$

where for the last equality we simply multiply out. The endresult is the Ridge estimator.