Problem Set # 6

Econ 722

1. This paper concerns the so-called "GMM-AIC" moment selection criterion from Andrews (1999) which takes the form

$$\text{GMM-AIC}(c) = J_T(c) - 2(|c| - p)$$

where $J_T(c)$ denotes the J-test statistic for the estimator based on the collection of moment restrictions indexed by c, $|c|$ denotes the number of moment conditions in specification c and p denotes the number of parameters. Characterize the asymptotic behavior of this criterion under the assumptions of the consistency theorem we proved in class.

NB: the next question is challenging but many of the steps of the argument you'll need to work out are sketched in the corresponding paper.

2. This question concerns a simple example of the FMSC from DiTraglia (2015). The idea is to choose between OLS and TSLS estimators of the effect β of a possibly endogenous regressor x. In particular, consider the following data generating process

$$\begin{aligned} y_{ni} &= \beta x_{ni} + \epsilon_{ni} \\ x_{ni} &= \mathbf{z}_{ni}'\boldsymbol{\pi} + v_{ni} \end{aligned}$$

where β and $\boldsymbol{\pi}$ are unknown constants, $\mathbf{z}_{ni}$ is a vector of exogenous and relevant instruments, x_{ni} is the endogenous regressor, y_{ni} is the outcome of interest, and ϵ_{ni}, v_{ni} are unobservable error terms. Let $(\mathbf{z}_{ni}, v_{ni}, \epsilon_{ni})$ be a triangular array of random variables that are iid *within each row of the array* such that $E[\mathbf{z}_{ni}\epsilon_{ni}] = \mathbf{0}$, $E[\mathbf{z}_{ni}v_{ni}] = \mathbf{0}$, and $E[\epsilon_{ni}v_{ni}] = \tau/\sqrt{n}$ for all n. You may assume that all random variables in this system are mean zero.

   (a) Let $\sigma_x^2 = \gamma^2 + \sigma_v^2$, and $\gamma^2 = \boldsymbol{\pi}'Q\boldsymbol{\pi}$ where $E[\mathbf{z}_{ni}\mathbf{z}_{ni}'] \to Q$, $E[v_{ni}^2] \to \sigma_v^2$, and $E[\epsilon_{ni}^2] \to \sigma_\epsilon^2$ as $n \to \infty$. Show that, under homoskedasticity and standard regularity conditions,

$$\left[\begin{array}{c}\sqrt{n}(\widehat{\beta}_{OLS}-\beta)\\ \sqrt{n}(\widetilde{\beta}_{TSLS}-\beta)\end{array}\right] \xrightarrow{d} N\left(\left[\begin{array}{c}\tau/\sigma_x^2\\ 0\end{array}\right],\ \sigma_\epsilon^2\left[\begin{array}{cc}1/\sigma_x^2 & 1/\sigma_x^2\\ 1/\sigma_x^2 & 1/\gamma^2\end{array}\right]\right)$$

where we define the OLS and TSLS estimators, as usual, by $\widehat{\beta}_{OLS} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}$ and $\widetilde{\beta}_{TSLS} = (\mathbf{x}'P_Z\mathbf{x})^{-1}\mathbf{x}'P_Z\mathbf{y}$ where $P_Z = Z(Z'Z)^{-1}Z'$. Hint: homoskedasticity will only need to hold *in the limit*. In particular, you'll need: $E[\epsilon_{ni}^2\mathbf{z}_{ni}\mathbf{z}_{ni}'] - E[\epsilon_{ni}^2]E[\mathbf{z}_{ni}\mathbf{z}_{ni}'] \to 0$, $E[\epsilon_i^2 v_{ni}\mathbf{z}_{ni}'] - E[\epsilon_{ni}^2]E[v_{ni}\mathbf{z}_{ni}'] \to 0$, $E[\epsilon_{ni}^2 v_{ni}^2] - E[\epsilon_{ni}^2]E[v_{ni}^2] \to 0$.

(b) Use the limit result from the preceding part to derive the FMSC for choosing between OLS and TSLS estimators for this problem.

(c) Show that, for this problem, FMSC is identical to the Durbin-Hausman-Wu test with a critical value of 2, which corresponds to $\alpha \approx 0.16$.

3. Consider a regression model of the form $y_t = \mathbf{x}_t'\beta + \epsilon_t$ where $\mathbf{x}_t$ is $(p \times 1)$ and satisfies $T^{-1}\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t' = \mathbf{I}_p$ and $\epsilon_t \overset{iid}{\sim} N(0,\sigma^2)$ where σ^2 is finite. You may treat the regressors as *fixed* rather than random in your derivations.

   (a) What is the MLE for β in this setting? Derive its finite-sample distribution. Is the MLE consistent for β?

   (b) Derive a closed-form expression for the Ridge Regression estimator of β in this setting, expressed in terms of the MLE and the shrinkage parameter λ. Use this result to write out the finite-sample distribution of the Ridge Estimator.

   Important Note: It will be helpful to factor a T from λ in your derivation. When you do this, you can still call the re-scaled shrinkage parameter λ rather than λ/T. Remember: we are free to choose λ so its precise scaling is irrelevant. Writing things this way will make your expression for the Ridge estimator match the (slightly different) example from the lecture notes and should help to avoid confusion in the parts that follow below.

   (c) First, suppose that we choose a fixed positive value for λ. Explain why the corresponding Ridge estimator will *not* be consistent for β as $T \to \infty$ in this case. Next suppose that, instead of fixing λ we decide to allow it to change with sample size. State sufficient conditions on the sequence λ_T to ensure that the Ridge estimator is consistent.

(d) Derive a closed-form expression for the LASSO estimator in this example, expressed in terms of the MLE and the shrinkage parameter λ.

Important Note: As in the Ridge example above, and for the same reason, if you encounter the term $T\lambda$ you can forget about the T and just call this λ.

(e) For sufficiently large λ, LASSO shrinks some coefficients all the way to zero and hence can be used to carry out variable selection. Derive an exact finite sample expression for the probability that LASSO decides to "exclude" regressor j, that is $P(\widehat{\beta}_j^{Lasso} = 0)$. Explain the intuition with the help of one or more plots.

(f) Now suppose we allow the LASSO shrinkage parameter to depend on sample size. Prove that $\lim_{T\to\infty} \lambda_T = 0$ implies that the probability of LASSO excluding a relevant regressor, i.e. one with a non-zero coefficient, converges to zero.

(g) Now consider the case of an *irrelevant regressor*, i.e. $\beta_j = 0$. What is the probability that LASSO excludes such a regressor? If we allow λ_T to depend on sample size, what condition on the *rate* at which $\lambda_T \to 0$ is required to ensure that LASSO excludes irrelevant regressors with probability approaching one in the limit? What happens if $\lambda_T \to 0$ at a slower rate?

(h) Combining the two preceding parts gives a *consistency* result for LASSO: provided that λ_T converges to zero at a sufficiently fast rate, all relevent regressors are selected and all irrelevant regressors excluded with probability approaching one in the limit. Crucially, however, this result depended on β_j being *fixed*. Suppose instead that we consider a sequence of local parameter values $\beta_{j,T} = \delta/\sqrt{T}$ where δ is a constant. When $\delta \neq 0$, this captures in asymptotic form the idea of "small but nonzero" coefficients. How do the results of the preceding parts change under these asymptotics? Discuss your findings.