ANSWER SHEET FOR PROBLEM 1 ON FINAL

(1) Suppose you are estimating the linear probability model

$$y_j = \alpha + \beta S_j + \varepsilon_j \,, \quad \boldsymbol{E}[\varepsilon_j \mid S_j] = 0 \,,$$

where S_j is an individual's score on a test and y_j is an indicator variable, 1 if the individual enters a four year college, 0 otherwise. The test scores have a possible range of 300 to 800, though in the sample at hand there are no scores at these bounds. Because the dependent variable can only be between zero and one and the equation will be used for predictions outside this sample, it must be that $0 < \alpha + 800\beta < 1$ and $1 > \alpha + 300\beta > 0$. This implies also that $-1 < 500\beta < 1$, taking the difference of the two inequalities, which both must hold. Assume the sample is large enough that the normal asymptotic theory for OLS and GLS applies. When the questions below have yes/no answers, explain your answer.

(a) [5 points] Is OLS unbiased in application to this model?
Yes, if we assume everything is i.i.d. and there is non-zero variance in test scores, the usual conditions for unbiasedness of OLS are met.

(b) [5 points] Suppose that instead of just using OLS uncritically, we had planned to use the OLS estimator only if it satisfies the inequalities and otherwise to discard it and use a probit model instead. If in the sample at hand the OLS estimator satisfies the inequalities, can we claim that our estimate is unbiased?
No. Unbiasedness is a frequentist concept. An estimator can't be unbiased in one kind of sample and not in another. Here it is clear that if the true β is large, we are more likely to get a $\hat{\beta}$ violating the upper bound than to get one violating the lower, so we will be discarding high estimates more than low estimates. This will make the estimator downward biased for high true values of β. Full credit for just realizing that because we are discarding the OLS estimator based on its size we don't have unbiasedness.

(c) [10 points] Explain why GLS improves on OLS for this model and how the GLS estimate would be constructed.
When y_j is either zero or one, with probability of a one $p = \alpha + \beta S_j$. The variance of y_j is then, as we discussed in class, $p \cdot (1 - p)$. It is therefore much smaller when p is close to zero or one than when p is close to .5. GLS should weight cross-product terms by the inverse of $p_j \cdot (1 - p_j)$, where $p_j = \alpha + \beta S_j$. This could be done either by "feasible GLS", first using OLS to get consistent estimates of α and β, then using those estimates to form the weights and re-estimating, or instead by using the full likelihood, with α and β appearing in the covariance matrix of disturbances as well as in the mean of the y_j's. In the latter case we would maximize the likelihood or form a posterior mean by MCMC simulation.

(d) [5 points] If we use GLS, but still plan to switch to probit if the estimates fail to satisfy the inequalities, will our estimator necessarily have higher variance than unmodified GLS, since GLS is the minimum variance linear unbiased estimator here?
No. GLS is only minimum variance in a restricted class of estimators. By discarding GLS estimates when they are impossible, we are probably making our estimator more accurate on average, and our estimator is not at all linear and not unbiased.

(e) [10 points] Suppose the GLS estimates $(\hat{\alpha}, \hat{\beta})$ do satisfy the inequalities, but the usual 95% confidence interval for β, $\hat{\beta} \pm 1.96\sigma_\beta$, has an upper end point greater than .002, and thus fails to satisfy $500\beta < 1$. If we just discard that part of the interval, truncating the interval on the right at .002, is it still a 95% confidence interval?

Yes. An interval with 95% coverage probability is one that contains the true parameter value with 95% probability for every possible true parameter value. By excluding from the interval parameter values that can't possibly be the true parameter value, we are not changing the coverage probability for any parameter value that could be the true value.

(f) [10 points] How would a Bayesian with a flat prior, relying on the asymptotic normality of the likelihood shape, construct a 95% posterior probability interval for β in this model? Use a sketch of the level curves of the likelihood and the constrained region of the parameter space to explain your answer. No arithmetic is required. [Hint: The four inequalities that α and β have to satisfy each require that (α, β) lie on one side of a straight line, and the four together restrict (α, β) to a parallelogram with vertexes the (α, β) pairs (0,0), (1,0), (-.6, .002), (1.6, -.002).]

The sketch should show circular or elliptical level curves of the approximately normal likelihood function and also the parallelogram. The question didn't specify whether the estimates were precise or not or even whether the likelihood peak was inside the parallelogram, so a variety of sketch shapes are possible. The interval is bounded by two horizontal (if β is the vertical axis) lines that between them contain 95% of the volume under the likelihood surface and within the polygon. Here is a precise computer version, much more precise than expected on the exam of course.