Solution Notes for MLE II Final Homework

Question 1: We may include a lag because we want to make inferences about conditional autocorrelation in the outcomes, or we may think that lags have to be included to remove bias in estimating the coefficients on x_{it} variables. We saw in the previous homework that FE with the lagged DV produced a significant coefficient on the lag. But, as shown in the notes, FE with a lag can be problematic. This motivates the Arellano-Bond estimator. Most people did fine in executing the Stata commands (`xtabond`), and most did okay in interpreting coefficients and reading off the tests of first and second order autocorrelation—e.g. appreciating that only the test of *second-order* autocorrelation is informative here and the test of first-order autocorrelation says nothing about the appropriateness of the estimator.

Many people had a hard time interpreting the Sargan test. The Sargan test is to see if overidentifying assumptions hold in instrumental variable estimation. Overidentification is not a problem to be corrected, but rather a situation that we *can use* in our estimation process. I am emphasizing this because many people wrote that the Sargan test was used to see if our data "suffered from overidentification," which is totally backward! Overidentifying assumptions are testable. Let me give some intuition for what is going on. Assume that (i) $y = \alpha + \beta x + \varepsilon$, (ii) $\text{Cov}(x, \varepsilon) \neq 0$ ("endogeneity"), but that (iii) $\text{E}(\varepsilon | z_1) = 0$ and $\text{E}(\varepsilon | z_2) = 0$ ("exclusion"), and that (iv) $\text{Cov}(x, z_1) \neq 0$ and $\text{Cov}(x, z_2) \neq 0$ ("valid first stage"). With these assumptions, we have two valid instruments, z_1 and z_2, which can be used to identify the coefficient, β, on x. Testing assumption (iv) is straighforward—just regress x on z_1 and z_2. (We can even do this when only one instrument is available.) Because our assumptions give us two valid instruments for identifying one parameter, β—that is, because β is overidentified—we can also test (iii). To test the exclusion assumption for z_2, we use z_1 as an instrument to identify β, and then test whether $\text{E}(\hat{\varepsilon} | z_2) = 0$; vice versa to test exclusion for z_1. This is along the lines of what the Sargan test is doing, but it does so in one step for all of the overidentifying assumptions. If the p-value of the test is less than our critical significance level, we reject the null hypothesis that the overidentifying assumptions are valid. (A few people were reading the test results as saying the opposite!) The distribution of the test statistic relies on the assumption that the errors are homoskedastic. I presume this is why Stata does not permit the Sargan test when `vce(robust)` option is used. The test may reject the null if (a) $\text{E}(\hat{\varepsilon} | z_k) \neq 0$ for some k or (b) if the homoskedasticity assumption is invalid. It's clear, then, that the Sargan test is testing more than just autocorrelation in the errors, although keep in mind that instances of (a) and (b) may be attributable to autocorrelation. Thus, no autocorrelation is necessary but not sufficient to pass this test with reasonable certainty. Arellano and Bond (1991) found that the Sargan test was not as sensitive to autocorrelation as was their autocorrelation test. The implication of course is that the two tests sometimes disagree, with the Sargan test being sensitive to other types of violations of assumptions, but also being less sensitive to particular violations associated with autocorrelation.

A brief note on the Arellano-Bond autocorrelation test: it is valid under many forms of dynamic panel model estimation, even though Arellano and Bond presented it in relation to two-step robust estimation. See the support documentation for `xtabond2` for a details on applying the autocorrelation test to a variety of dynamic models.[1] Many people conducted a Sargan test after one-step estimation with `vce(gmm)` (assumption of homoskedasticity) and compared these results to the autocorrelation test after two-step estimation with `vce(robust)`. The two are not expected to agree. The latter uses the robust variance-covariance to normalize the test statistic; an autocorrelation test under a homoskedasticity assumption would use a homoskedastic variance covariance matrix. Again, see the support documentation for `xtabond2` (pp. 34-5).

[1] Roodman D. 2006. "How to Do xtabond2: An Introduction to 'Difference' and 'System' GMM in Stata." *Center for Global Development Working Paper*, Number 103.

Those who were able to obtain both the Sargan test and autocorrelation test after one-step homoskedastic estimation found that the two were somewhat less discordant, but not always in agreement. It seems that there were limits in some, but not all, versions of Stata's `xtabond` that made it difficult to perform all tests on the same model.

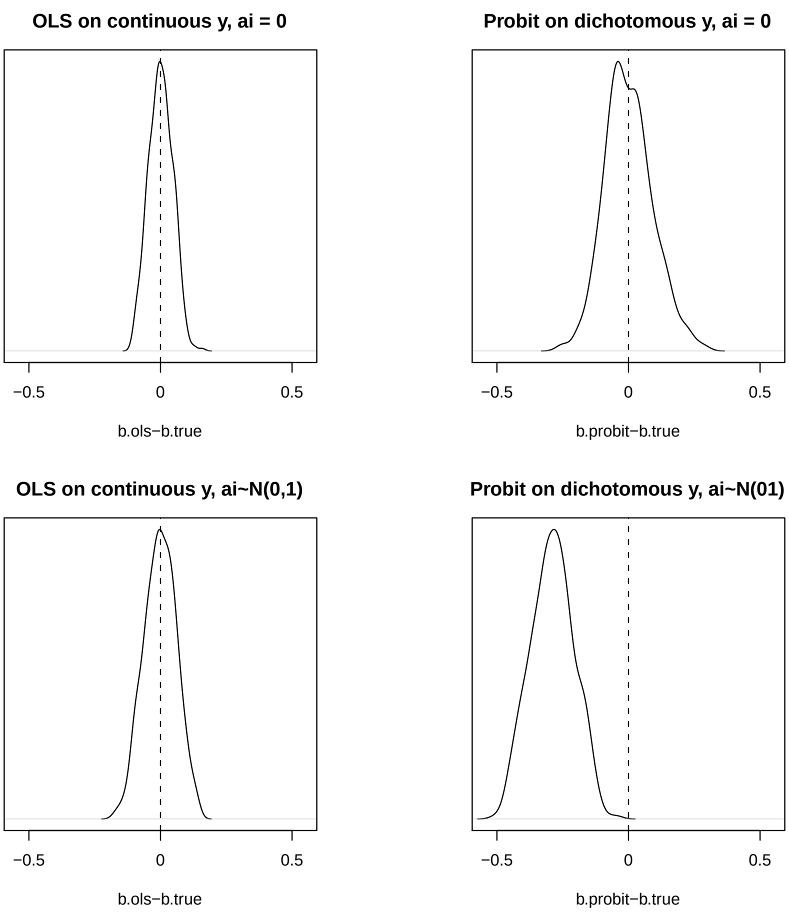
Putting this all together, a reasonable approach might be to use robust standard error estimation to deal with heteroskedasticity (and thus rendering the Sargan test baseless), and then test for remaining autocorrelation using the autocorrelation test (which is more sensitive to such problems than the Sargan test anyway).

Question 2: People were generally fine in implementing the methods for this. Many people still seem to limit themselves to only interpreting the signs and significance of coefficients, without exploring their substantive implications. As mentioned in the previous solutions note, you should be comfortable with substantive interpretations derived from logit and probit coefficients.

Few people really picked up on the following points about unit effects in limited dependent variable models. I think these points are important, so let me expand on them a bit: When y_{it} is a continuous variable with unbounded support (e.g., not constrained to be greater than 0, etc.), random effects are more of an efficiency problem than a bias problem. If α_i is orthogonal to explanatory variables x_{it}, then we may have a violation of homoskedasticity. But this does not create problems of bias for estimating the coefficients on x_{it} via OLS, which ignores the α_i's. However, this logic does not carry through when y_{it} is a limited dependent variable—e.g., dichotomous. The reason is the non-linearity of α_i's and x_{it}'s relationship to y_{it}. Even with orthogonal α_i's, estimation of coefficients on x_{it} in a manner that ignores the α_i's will be biased. Thus, random effects probit is motivated by an interest in bias minimization, whereas in the OLS world, random effects were just an efficiency concern. The different consequences of random effects for OLS and ordinary probit are illustrated in the figure below. The left panels show bias in 500 simulated OLS estimates of β_1 for data generated as $y_{it} = .5 + x_{it} + \alpha_i + \varepsilon_{it}$. The right panels show bias in 500 ordinary probit estimates of β for data generated as $y_{it} = \text{Bernoulli}(\Phi(.5 + x_{it} + \alpha_i + \varepsilon_{it}))$. In all four simulations, $x_{it} \sim \mathrm{N}(0,1)$, and $\varepsilon_{it} \sim \mathrm{N}(0,1)$. In the top two simulations, $\alpha_i = 0$ (i.e. there are no unit effects at all); in the bottom two, $\alpha_i \sim \mathrm{N}(0,1)$ (i.e. orthogonal unit effects). We see that exogenous unit effects have only efficiency consequences for OLS, but affects both efficiency and bias for ordinary probit. The bias here is a type of attenuation bias.

Fixed effects models go one step further by allowing us to relax the assumption of orthogonal α_i's. Of course, this requires that we also accept that we cannot identify the partial effects of time-invariant variables. Our ability to identify the partial effects of slow-moving variables is also compromised. Finally, in the conditional MLE, fixed effects are handled by considering only cross-sections with variation in the outcome variable, limiting the set of observations with which we can work. Fixed effects estimation is thus a rather conservative framework, but for some that is what makes it appealing. Finally, it should be understood that fixed effects logit estimation purges components of x_{it} that are fixed over time. These components cancel out in the same way that the α_i's cancel out. (To check this, write out the following decomposition and plug it into the likelihood equation in the notes: $x_{it} = \mu_i + \nu_t + \zeta_{it}$.) An implication is that any stable, additively separable effects of α_i on x_{it} are thus swept out of the analysis. This is how fixed effects allows us to identify coefficients on x_{it} even when there is correlation with α_i.

Figure 1: **Magnitude of bias due to random effects for OLS and ordinary probit**



The top two figures report results of 500 simulations with no random effects; the bottom two report results of 500 simulations with random effects that are orthogonal to the covariates.