Comments on the memos

Monique Borgerhoff Mulder and Sam Bowles
12 February 2007

The memos are really great! They are full of interesting estimates, interpretations, and ideas about next steps. Below are some ideas stimulated by reading the memos.

1. It will be useful to produce a large matrix of groups (rows: Krummheorn, Kipsigis, etc) x types of wealth (columns: land, cattle, somatic, RS) with entries of the best summary statistic (β estimated from the most comparable possible specifications, as well as a measure of wealth inequality in the two generations). We have started doing this, and will finalize at the workshop with all your input.

2. Selection bias is a big item to which we should direct serious attention. There are many issues here, but two jump to mind: first cases in which a primogeniture system results in our having data on dad's wealth and first son's wealth but not on the other sons. Second, and related, is bias arising from sampling procedures that lose emigrating offspring or offspring for whom wealth is not recorded. Jan's comments on 'celibate sons' and emigrating sons are pertinent here: they impart an upward bias to the estimates of β and ρ. Greg introduces some neat ways of trying to gauge the effect of these biases.

3. Inequality, transmission and disequilibrium states. In some data sets we have very equal wealth distribution but high levels of transmission (e.g. the Abosi sample of Monique's compared to Jan's Krummheorn data set both show high β;s but only the former latter show substantial wealth inequality.) The Krummheorn data may reflect a long term equilibrium, the Abosi data not). We need to think about each of these systems with respect to equilibrium dynamics.

4. Significance levels: we are at a very preliminary stage and do not want to be throwing out estimates that fail the normally stringent significance tests. I suggest we play close attention to the size of (unit free) measures like ρ and β and adopt high tolerance levels (for the time being) about significance levels.

5. Is the degree of transmission given by the nature of the wealth or by the social institutions of the group (or other influences unrelated to the nature of the wealth)? Why not pool all of our data for wealth of a given kind and test for differences among our groups (standard F tests, etc). We could do this for livestock, land, and RS (I think). We would also need to think about how we

might test the alternative hypothesis (institutions) and how we would code these;
a simple first stab would be good to look at whether intergenerational correlations are better explained by "resource type" or "institutions, norms, and other group-specific influences".

6. A number of memos have pointed out that (random) errors in wealth measurement may significantly downward bias estimates of both β and ρ. Some (Greg) have addressed this. In other cases we have multiple measures of the same thing. For example, Mike could average the somatic wealth measures taken over many years (thereby reducing the noise to signal ratio). Rob could produce a direct measure of the correlation between the true and observed measures of land ownership (assuming uncorrelated errors in the two sources of information, this is just the square root of the correlation between the two sources). Using his data to take account of measurement error increases the father son correlation in land holdings from 0.26 to 0.37, so the effects can be pretty substantial (recall that the measurement error for males was comparatively minor). It could be that all of the difference between men and women in Rob's calculations is due to the different levels of measurement error. Monique's Pimbwe data set apparently has 6 years of data collection in which the same questions were asked (of the same people); this would allow a measure of error, or at least the opportunity to average out the error. We need to see if we can make any headway on this problem with all of data sets.

7. Where we have age (A) and other interactions with wealth we need comparable calculations of the derivative of offspring wealth with respect to parent wealth (not just the directly estimated coefficients).

8. The logarithm of zero problem. I'm not sure everyone handled this in (even approximately) the same way. Suppose for some observations the parental wealth $W = 0$ Adding 1 to all the observations is ok and will not produce much of a bias if the mean of W is large, but where it is small (e.g. Rob's data) it would be a good idea to inflate the W measure so that the measure used in the calculation is $xW + 1$ e.g. $100W + 1$, thereby insuring that the arbitrary 1 has little influence on the regressions. This is what Rob did except that he did not add the 1 to all individuals only those with zero land.

9. Sibling correlations. We need to standardize how we compute the correlations. I think that the method in the Bjorklund et al paper is the right way to go.

10. Much of our data concern men only, but as Mary points out (p. 8 of her memo) in many societies the transmission process differs greatly by sex, and we should think about these differences and how to empirically capture (and understand)

them. This also concerns Mary's point about the appropriate aggregate of parents traits (summing education may make sense, while summing income may not, e.g. in Bangalore where most of the women do not work for pay). Mary's list of next steps bears reading.

11. There are a number of odd findings that jump off the page (the low β that Monique found for Kipsigis II, the very high father son transmission of surviving children that Mike found). Obviously we should look at these and other findings with special care.

12. Restricted range bias. As Greg points out, where our sample is biased so that the parent's wealth data come from a restricted range of the parents true generation, the resulting correlations (and β's) will be downward biased. Greg has addressed this more or less mechanically. It probably comes up in other data sets. There are standard ways of correcting for restricted range bias.

13. In some of the smaller data sets the estimates may be quite sensitive to a few observations (Bobbi presents a striking case of this, Mr. 70625). We may want to do a systematic sensitivity analysis (the standard method is to eliminate the most influential observation, the top two, the top three, etc).

14. We need to think carefully about the consequences of selecting one (or a few) sources of wealth. If cows and acres in Monique's case, or bay trees, fishing and other economic pursuits (in Rob's case) were substitutes of each other, low β's for specific resource types might not accurately capture high β's for total wealth. In cases where major wealth types are measured this can be dealt with by finding a common value for summing wealth types (e.g. cash value); in other cases, it will be important to analyse all major sources of wealth.

15. Related to point 2, we need to think more about the role of siblings. A number of memos pointed out that it might be necessary to control for sibling number, and some did. If there is variation in wealth among siblings, entering siblings into the model will reduce β's. Whether we keep siblings in or out of the model depends on whether we are looking at intergenerational correlations from parents or offspring perspective (no?), but whatever we do we need to be consistent across studies.

16. Many have mentioned additional data that could be deployed to shed light on these questions (some is collected but not yet analyzed, some would have to be collected). We need to catalogue all of the additional data that may used (possibly in conjunction with the monster matrix suggested in point 1).

17. We also need to brainstorm about others who should be recruited to the project. We suspect that there may be quite a few who have or may have data

that would be relevant. So come with a list of other studies where we should be looking to find such estimates, or whose authors could be convinced into analysing their data for us

18. We encourage those of you who have not been able to include age in your calculations to explore your data (other samples from your data that can convince you/us that age is not a factor that is likely to affect beta estimates. (This is particularly relevant for a study like Bill's or Greg's or possibly Richard's)