systematically correlated with the explanatory variables. Farms in low-quality marginal areas (semi-deserts) are typically large, and farms in high-quality land areas are often much smaller. That a garden adds more value-added per hectare than a sheep station does not imply that sheep-stations should be reorganized as gardens. The omitted quality variable is negatively correlated with A, and so causes the estimated coefficient to be downward biased, from the true value of zero to the observed negative value. Indeed there is some evidence that controlling for quality either reduces or removes the effect [see Bhalla and Roy (1988) and Benjamin (1993)).

Similar arguments apply to the other variables in the production function. For example, it is sometimes found that the returns to fertilizer use, estimated from regression coefficients, are many times larger than would be consistent with productive efficiency [see for example Benjamin and Deaton (1988) for Cate d'Ivoire and Bevan, Collier, and Gunning (1989) for Kenya and Tanzania]. Should fertilizer use be encouraged, and extension services expanded? Not if what we are seeing is that the farms with the higher quality land, or with the most go-ahead farmers, are also those who adopt new technologies. Output is high, not because of the return to inputs, but because of unobservables, land and farmer quality, that are correlated both with inputs and outputs.

Omitted heterogeneity induces correlations between explanatory variables and the error term in a way that has the same consequences as simultaneity bias. Indeed, the production function is likely to suffer from genuine simultaneity bias even in the absence of heterogeneity; inputs, like outputs, are under the control of the farmer, and can have no general claim to exogeneity. The combination of genuine simultaneity and heterogeneity has the further effect of ruling out the use of lags to remove the former; while it is true that seeds have to be planted before the crop is harvested, heterogeneity across farmers will mean that seeds are not exogenous for the harvest, a problem that I shall return to in Section 2.2 in the context of using predetermined variables with panel data. The result of all these considerations is that the regression function of physical output conditional on physical inputs will rarely be informative about the underlying technology.

There are a number of possible econometric solutions to these problems. Note first that, under the standard neoclassical assumptions of the farm-household model, the appropriate exogenous variables for production are not inputs, but the prices of inputs, and the appropriate estimation technique is either instrumental variables applied to the physical production function, or the estimation of a dual specification, in which the technology is specified as a profit function whose derivatives are the demand functions for inputs and the supply functions of outputs, all functions of prices.

There are two problems here, one theoretical and one practical. First, many