1979-82 and this contaminates Pn as a proxy for social learning opportunities. Such effects do not, however, contaminate Pu: permanent idiosyncracies of the cluster will be picked up in Pg, temporary idiosyncracies pertinent to the decision period will be picked up in Pn. This becomes important in view of our results.

We introduce heterogeneity through the gender of the household head. Around a quarter of the households in our sample are female-headed. The survey permits a distinction between the head and the decision taker. In some female-headed households major agricultural decisions are taken by non-resident males, and in these cases we classify by the gender of the decision taker. As with the sequence of adoption, heterogeneity can only influence social learning to the extent that it is readily observable, and the gender of the head clearly meets this criterion. Female-headed households may differ in a number of economic respects from male-headed households, for example, being less able to borrow. The gender of the head therefore proxies a set of less observable economic dimensions of heterogeneity. In our application we test whether there is endogenous restriction of the window of observation by gender. However, it should be stressed that we are unable to distinguish between two interpretations of the test. The more obvious (and in our view more likely) interpretation is that agents are themselves observing the gender of the household head as a low-observation-cost proxy for the underlying economic characteristics. Alternatively, agents may be observing these economic characteristics directly and restricting their window of observation on these criteria, with gender simply being the low-observation-cost variable which the researcher uses to proxy them. Since our test is concerned not with gender per se, but with whether there is endogenous restriction of the window of observation, this distinction is unimportant, but it does caution against the interpretation of our results for other purposes.

4. Econometric Specification

We use logits to explain two sets of behaviour, innovation in 1979-82 and in 1976-79, in terms of social learning (Pg, Pu, Pn) and endowments (x_i). Let the latent (unobservable) variable y^*_i measure the attractiveness to farmer i of coffee adoption: adoption takes place if y^*_i > 0. The latent variable is a function of the behaviour of other households and of the endowments of the decision taker and his household. The decision taker can observe the
decisions of his neighbours but not the corresponding latent variables. The adoption decision is determined by the value of \( y^*_i \) which is the expected value of an underlying random variable, \( e_i \).

\[
y^*_i = E_i[e_i]
\]

where the subscript \( i \) is added to both the expectations operator and the underlying variable \( e \) to indicate the individuality of the process. The distribution of \( e_i \), as seen by \( i \), is affected by \( j \)’s (all \( j \neq i \)) decision to adopt. In the above terms, \( j \)’s adoption means that \( y^*_j > 0 \). If \( e_i \) and \( y^*_j \) are correlated, with correlation coefficient \( r_{ij} \), then the expected value of \( e_i \), conditional on \( y^*_j > 0 \), becomes:

\[
E_i[e_i | y^*_j > 0] = E_i[e_i] + r_{ij}s_i f_j(0)/(1-F_j(0)),
\]  
(1)

where \( s_i \) is the standard deviation of \( e_i \), \( f_j \) the density function of \( y^*_j \) and \( F_j \) the distribution function of \( y^*_j \), as seen by individual \( i \).\(^5\) Equation (1) applies if both distributions are normal. Suppose now that the formula approximately holds for our logit case, in which \( y^*_j \) is assumed to follow the logistic distribution. Then the ratio of \( f_j \) to (1-\( F_j \)) is equal to \( F_j \), and \( F_j(0) \) is \( j \)’s probability of not adopting.

We have distinguished three components of social learning: the popularity of intentions to innovate, \( Pn \), the popularity of pre-gestation innovations, \( Pu \), and the popularity of gestated innovations, \( Pg \). For innovation in 1979-82 we have proxies for each of these, while for innovation during 1976-79 we have a proxy only for \( Pn \) and a composite \( Pg+u \). The weight on any of the resulting four measures of popularity can be estimated as in (2) below, with one exception:

\[
y^*_i = a'x_i + b.w_i'y + u_i
\]  
(2)

where \( y \) is a vector of which element \( j \) takes the value 1 if farm \( j \) belongs to the innovating group (through an intention to innovate, a pre-gestation investment, a gestated investment or

---

\(^5\)The formulas for these expectations with truncated distributions are given in Maddala (1983).
either type of investment respectively) and 0 otherwise, w is i’s weighting vector on that type of popularity, a is a vector of parameters and u is the random term. The exception is Pn, for although Pn appears to be exogenous to the household whose behaviour is being analysed since it measures the intentions only of other households, it is in fact endogenous. Household i’s intention depends upon what other farmers intend to do, but, as the same model applies to the other households, the decisions of others depend in part upon whether household i intends to innovate. Hence, the inclusion of Pn as directly observed makes the model inconsistent: the latent variable of farm i is indirectly related to whether farm i itself intends to adopt. This precludes a model of the simple form of (2). However, if (2) is interpreted as the conditional latent variable, the model can, in principle, be estimated once the conditional probabilities are transformed into the joint probability of the whole group.

If p_i is the probability of adoption by i during the period, we have

\[ \log \text{odds-ratio} \ p_{ii} = a'x_i + by_i \]

and

\[ \log \text{odds-ratio} \ p_{ji} = a'x_j + by_i \]

which is consistent with the log-linear model

\[ \log \Pr(y_i, y_j) = a_0 + a'x_i y_i + a'y_i y_j + by_i y_j \]  \hspace{1cm} (3)

where a_0 is such that the four probabilities sum to 1. A necessary assumption for (3) to hold is that the error terms in the two latent variables of i and j are independent.

The model can be extended to all households in the window of observation. In principle it is possible for the log-linear model to distinguish separate coefficients b_i for

---

*This is one of the "logically inconsistent" models discussed by Maddala (1983, p. 119). A further discussion of this type of simultaneous model, which is possible for continuous variables but not for discrete variables, is given by Cramer (1986, p. 181).

7 We would like to thank Chris Elbers for this suggestion.

*See Maddala (1983).
every pair (i,j), and higher order interaction terms like \( b_{ik} \) for triples (i,j,k). If we assume that all these interaction terms are equal to \( b \), then for the vector \( y = (y_i) \), where \( y_i = 1 \) if i adopts and \( y_i = 0 \) if not, i= 1,...n:

\[
\log \Pr(y) = a_0 + y'Xa + bm(m^2-1)/6
\]  

(4)

where \( X \) is the matrix of the endowment variables for all observations, and \( m \) is the number of adopters in the group. If only first-order interaction terms are included, the last term changes to \( bm(m-1)/2 \). The constant \( a_0 \) is such that the sum of the probabilities of all possible configurations of \( y \) equals unity. Equation (4) shows that the probabilities of each \( y_i \) being 1, conditional on the number of adopters (m) in the group, are now independent, because the last term of (4) is constant given the condition, thus eliminating the endogeneity problem.

From this model no simple expression can be derived for the marginal probabilities of household i. Hence estimation should be based on a groupwise approach. If the window of observation is the entire cluster, so that there is only one group, the likelihood to be maximized is given by (4), where we should take the values of the vector \( y \) as observed in the sample. If there is more than one group, the likelihood to be maximized is the sum over groups of expressions like (4).

Finally, we take into account that we do not know the endogenous window of observation, that is, the chosen reference group of the household, nor do we have all relevant observations in our sample. Hence, we can do no more than assume that the observations in the cluster form a basis for an estimate that is relevant for the farmer. Expected adoption is the estimate for the adoption in agent i's chosen window of observation. This gives:

\[
y^*_i = a'x_i + bw_i'E[y] + u_i.
\]  

(5)

There are two differences between (4) and (5): the second term in (4) has been replaced by a non-random expression and the only interaction terms are those involving agent i. If we ignore the possibility of idiosyncratic expectations then the expected value of adoption by others is precisely the probability of adoption, so that the model becomes:
\[ y^*_i = a'x_i + b.w_i'p + u_i, \]  

(6)

where \( p \) is the vector of probabilities of adoption.

This model is estimated iteratively in the next Section. In the first iteration \( b \) is set equal to zero; the resulting estimates of the vector \( a \) are used to calculate \( w_i'p \) which is used in the second round to recalculate the parameters. In the next stage the estimated \( Pn \) is added to the explanatory variables in the logit. Expected probabilities are then calculated until the estimated coefficients converge.

In the spirit of the EF and B models, this model is parsimonious in its assumptions on the information possessed by an agent on other agents: agents observe only an intention to innovate and the presence of gestated and ungestated investments. Since social learning is only rational in the context of costly information, such parsimony is appropriate. This can be contrasted with the implicit information assumptions of the econometric specification of Case (1992). In her model of innovation in a peasant economy the social learning variables are the latent variables \( (y^*_j, j \neq i) \) of the other households:

\[ y^*_i = a'x_i + b.w_i'y^* + u_i. \]  

(7)

Solving for \( y^* \) gives:

\[ y^* = (I-bW)'Xa + (I-bW)'u \]  

(8)

where \( y^* \) and \( u \) are vectors, and \( W \) and \( X \) are matrices. The implicit assumptions are extreme: farmers fully know the circumstances of their neighbours, even including the error terms. While the farmer may, at a cost, know the household characteristics of his neighbours, it is implausible that he would also have all the information which the researcher

---

*This procedure is repeated until convergence, i.e. until the coefficient of the copying variable changes by less than 0.001. The introduction of the estimated variable among the explanatory variables in the logit estimation makes the standard errors of the parameters biased when estimated in the usual way. Adjusted standard errors are discussed in the Appendix.*

*Clearly, the error term in (8) is heteroskedastic and the estimation procedure must allow for this.*

15
includes in the error term: this would imply that all farmers can predict their neighbours' adoption behaviour perfectly.\textsuperscript{11}

5. Results

The model explains coffee adoption during the periods 1979-82 and 1976-79 in terms of the proxy measures of the three types of popularity which are postulated to induce social learning, and the endowments which determine the suitability of the household for coffee adoption. The explanatory variables used in the logit are:

\textit{Popularity}

Post-gestation popularity (Pg) the proportion of farmers in the cluster who were gestated coffee

Pre-gestation popularity (Pu) the proportion of those farmers in the cluster not growing gestated coffee who had already planted ungestated coffee.

Composite popularity (Pu+g) the proportion of farmers in the cluster already growing coffee

Intention popularity (Pn) the expected value of adoption by others in the same cluster (see text)

\begin{footnotesize}
\textsuperscript{11} A variant on the Case model which we have investigated is to relax the extreme assumption about knowledge of error terms, assuming instead that farmers only know the deterministic part of the neighbours' latent variable. Hence:

$$ y^*_i = a'x_i + b'w_i'E[y'] + u, $$

where

$$ E[y'] = y' - u. $$

The reduced form for this model is

$$ y^* = (1-hW)'Xa + u, $$

which has a homoskedastic error term. While this model relaxes Case's implicit assumption of full knowledge of other farmers' latent variables, it still implies that farmers know all relevant (but in principle observable) characteristics of other farms. Estimation results of this model are not reported but are available from the authors.
\end{footnotesize}
Endowments:

age  
age of the decision taker

education  
a dummy variable taking the value 1 if the decision taker has at least some primary education

adults  
the number of household members over the age of 15

area  
the area of the household’s arable land

province  
a dummy variable, 1 for Central Province, 0 for Nyanza Province

finance  
access to credit, measured by a dummy variable taking the value 1 if in 1975 the household had a bank account or the decision taker had a wage job in that year

gender  
a dummy variable taking the value 1 if the decision taker is female

The means and variances of the explanatory variables are shown in Table 1, separately for male and female decision takers and for adopting and non-adopting households. The Table shows that for male decision takers there is little difference in endowments between innovators and non-innovators. By contrast, among female decision takers some appear to matter, notably age and education. To allow for a non-linear effect of age the quadratic term age^2 is included; and to allow for interaction between the effects of gender and education we include the term education*gender.

The inclusion of constant terms in both logits subsumes all temporally specific variables such as the price of coffee. Since the model seeks to explain only two periods of adoption there is no scope for a richer specification of time-dependent variables. For example, the price of coffee was higher during 1976-79 than during 1979-82 and so other things being equal, we would expect more adoption and hence a larger constant term.

We impose the restriction that coefficients are common to the two periods in all cases where such a restriction is not rejected by a likelihood ratio test (at 5% significance level). This applies to most of the endowment explanatory variables: female, education, education*female, adults and finance. Although the proportion of households growing coffee pre-1975 appears in both logits, it does not proxy the same concept and so we would not expect it to have the same effect. Recall that in the explanation of adoption during 1979-82
it proxies the popularity of gestated coffee, whereas in that for 1976-79 it proxies both gestated and ungestated coffee growing, thereby losing implicit information about the past history of output. It would therefore be incorrect to impose equality of the coefficients and indeed such a restriction is rejected by the likelihood ratio test.

The results (Table 2) reveal the weights on three of the four measures of popularity to be positive and significant and support the hypothesis that the window of observation is endogenously restricted. The post-gestation popularity and pre-gestation popularity variables are both significant, as is the composite measure Pu+g. The post-gestation popularity variable may, as noted in Section 3, pick up fixed effects, but this is not the case for pre-gestation popularity. That the latter variable is significant can therefore be interpreted as clear evidence for social learning. There is no significant weighting on intention popularity: although the coefficient is positive the t-statistic is only 0.3. Recall that even in the presence of social learning, the sign of the weighting on intention popularity is ambiguous because it constitutes a reason for delaying innovation in order to benefit from subsequent social learning opportunities. Social learning is gender-specific. The weighting on pre-gestation popularity is positive and significant for decision takers of the same sex but insignificant for those of the opposite sex. This effect is symmetrical: males copy males and females copy females. This endogenous restriction of the window of observation does not, however, extend to the weighting on post-gestation popularity: when this variable is disaggregated the gender distinction is not significant.

The variables measuring the household's endowments are not the focus of this paper, however, one result is highly pertinent. The coefficient on the gender of the decision taker is negative and significant: when decisions are taken by a female head of household then adoption is less likely. Indeed, this is the most significant of the endowment variables. This suggests that if agents choose an easily observable characteristic on the basis of which they restrict their window of observation, gender is the most likely candidate. However, gender presumably proxies a set of less observable characteristics which would reveal the origins of female disadvantage. As discussed in Section 2, although there is clear evidence for endogenous restriction of the window of observation, the precise nature of that restriction is open to two interpretations. Either gender is an observable proxy used by agents in window restriction, or agents are directly observing the underlying characteristics with gender merely serving as the proxy for them observable by the researcher. Fortunately, in the present
context this observational equivalence is unimportant since the issue has been whether there is endogenous restriction of the window rather than the nature of that restriction.

6. Conclusion

If information on the consequences of innovation is costly, agents may have an incentive to learn from the behaviour of others. Ellison and Fudenberg (1993) have proposed a model in which the agent positively weights the popularity of an innovation with other agents in their own decision, and have suggested that it might be rational for agents to choose to restrict their observation of others to a proper subset with similar characteristics. We have extended the model to allow for three distinct types of popularity weighting derived from the observable investment sequence of intention to invest, pre-gestation investment and post-gestation investment and have tested for such social learning in the coffee adoption decisions of Kenyan peasant households. We have demonstrated that these agents placed considerable weight upon the two latter measures of popularity, and that there was endogenous restriction of the window of observation according to gender: decision-takers were only influenced by the popularity of innovation among their own gender.

In our analysis we made no allowance for differences in the cost or quality of the information gained from social learning. However, the three types of popularity weighting which we have investigated are not equally easy to observe nor do they yield equivalent information. The observation of intentions is arguably costly and its value limited by their tendency to change. By contrast, ungestated investments are highly visible in a small community and reflect near-current information possessed by the innovator (but not directly observable). Post-gestation investment is less valuable in that it reflects an undifferentiated mass of more and less remote information, but it does generate a potentially observable output signal. Possibly, our result that there is endogenous window restriction for pre-gestated investment but not for that post-gestation indicates that when agents are free riding on those decisions of others which have no observable consequences they place greatest weight on agents similar to themselves, whereas when they can supplement popularity weighting with observation of consequences the window is widened to gain knowledge of output variation. The incorporation of such considerations may be a promising extension of the analysis of social learning.
References


Table 1: Means and Variances of Variables

<table>
<thead>
<tr>
<th></th>
<th>male decision takers</th>
<th></th>
<th>female decision takers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-adopting</td>
<td>adopting</td>
<td>non-adopting</td>
<td>adopting</td>
</tr>
<tr>
<td></td>
<td>N = 125</td>
<td>N = 62</td>
<td>N = 48</td>
<td>N = 6</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>variance</td>
<td>mean</td>
<td>variance</td>
</tr>
<tr>
<td>age</td>
<td>46.76</td>
<td>290.02</td>
<td>48.35</td>
<td>212.92</td>
</tr>
<tr>
<td>education</td>
<td>0.64</td>
<td>0.23</td>
<td>0.56</td>
<td>0.25</td>
</tr>
<tr>
<td>adults</td>
<td>3.31</td>
<td>3.15</td>
<td>3.65</td>
<td>3.87</td>
</tr>
<tr>
<td>area</td>
<td>2.95</td>
<td>5.47</td>
<td>3.72</td>
<td>11.42</td>
</tr>
<tr>
<td>province</td>
<td>0.42</td>
<td>0.25</td>
<td>0.60</td>
<td>0.24</td>
</tr>
<tr>
<td>finance</td>
<td>0.23</td>
<td>0.18</td>
<td>0.24</td>
<td>0.18</td>
</tr>
<tr>
<td>gestated investment</td>
<td>0.18</td>
<td>0.05</td>
<td>0.38</td>
<td>0.08</td>
</tr>
</tbody>
</table>
Table 2: Estimation Results

<table>
<thead>
<tr>
<th></th>
<th>Innovation 1976-79</th>
<th>Common Coefficients</th>
<th>Innovation 1979-82</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>t</td>
<td>t</td>
</tr>
<tr>
<td>constant</td>
<td>-8.9 3.6</td>
<td></td>
<td>-5.2 2.3</td>
</tr>
<tr>
<td><strong>popularity variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>post-gestation popularity</td>
<td>2.54 1.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-gestation popularity:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>same sex</td>
<td>3.6 2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>opposite sex</td>
<td>1.6 0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>composite popularity</td>
<td>3.28 3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>intention popularity</td>
<td></td>
<td>0.60 0.3</td>
<td></td>
</tr>
<tr>
<td><strong>endowment variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>province</td>
<td>0.54 1.3</td>
<td>1.85 2.9</td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td>-3.4 3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>0.21 2.4</td>
<td>0.058 0.6</td>
<td></td>
</tr>
<tr>
<td>age.age</td>
<td>-.0017 2.2</td>
<td>-.0009 1.0</td>
<td></td>
</tr>
<tr>
<td>education</td>
<td>0.10 0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>education.gender</td>
<td>2.49 2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>adults</td>
<td>0.15 1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>area</td>
<td>0.014 0.2</td>
<td>0.17 2.0</td>
<td></td>
</tr>
<tr>
<td>finance</td>
<td>-0.82 2.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**performance**

<table>
<thead>
<tr>
<th>Actual adoption</th>
<th>Predicted</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no  yes</td>
<td>no  yes</td>
</tr>
<tr>
<td>no</td>
<td>196 6</td>
<td>171 2</td>
</tr>
<tr>
<td>yes</td>
<td>33 6</td>
<td>17 12</td>
</tr>
</tbody>
</table>
Appendix: Derivation of the Variance-Covariance Matrix when estimated explanatory variables are included.

The model is for each i:

\[ y_i = x_i'a + b.w_i'p, \quad i = 1, ..., n \]

where \( y \) is the \( n \times 1 \) latent variable, \( x_i \) contains the \( k \) explanatory variables of \( i \), \( a \) is a \( k \times 1 \) vector of corresponding parameters, \( w_i'p \) is the weighting of popularity, with \( w_i \) containing the \( n \) weights of the farms as seen by \( i \), \( b \) is a scalar, and \( p \) is an \( (n \times 1) \) vector, defined by

\[
p_i = \frac{\exp(y_i) / [1 + \exp(y_i)]}{1 + \exp(y_i)}
\]

Let \( q_i \) be defined by

\[
q_i = -\log(1-p_i)
\]

and \( d_i \) by

\[
d_i = 1 \text{ if farm } i \text{ has adopted; } = 0 \text{ if farm } i \text{ has not yet adopted.}
\]

The log-likelihood of the observations in the sample is

\[
\log L = d'(Xa + b.Wp) - s'q
\]

where \( s \) is a \( n \times 1 \) vector of ones, and \( X \) is the \( n \times k \) matrix of explanatory variables.

The first derivative of \( p \) with respect to \( a \) is a \( n \times k \) matrix

\[
P_p = A(X + b.Wp),
\]

where \( A \) is a diagonal matrix with elements \( p_i(1-p_i) \); \( W \) is a matrix with rows consisting of the weighting vectors \( (w_i) \) of each farm. Thus

\[
P_p = (I - b.AW)'AX
\]

Similarly, the first derivative of \( p \) with respect to \( b \) is a \( n \times 1 \) vector

\[
p_b = (I - b.AW)'AWp
\]

The first derivative of \( \log L \) with respect to \( a \) is the \( 1 \times k \) vector

\[
d\log L/da = (X + b.Wp)'(d-p)
\]

The first derivative of \( \log L \) with respect to \( b \) is a scalar
\[
d\log L/db = (p + b \cdot p_b)'W(d-p)
\]

In the derivation of the second-order derivatives we have ignored the derivatives of \( P_a \) with respect to \( a \), of \( P_s \) with respect to \( b \), and of \( p_b \) with respect to \( a \) and \( b \).

The matrix of second-order derivatives of \( \log L \) with respect to \( a \) and \( b \) (which is equal to minus the inverse of the V-C Matrix) is therefore

\[
VC^{-1} = \begin{pmatrix} K & k \\ k' & m \end{pmatrix}
\]

where \( K \) is the \( k \times k \) matrix of second-order derivatives with respect to \( a \), \( k \) the \( k \times 1 \) vector of cross-derivatives, and \( m \) the (scalar) second-order derivative with respect to \( b \),

\[
K = -(X + b WP_a)'P_a \\
K = -(X + b WP_a)'p_b + (d-p)'WP_a \\
m = -(p + b p_b)'WP_b + 2(d-p)'WP_b
\]