Appendix 2. Supplementary text on methods for Garrett et al. "Explaining the persistence of low income and environmentally degrading land uses in the Brazilian Amazon."

A2.1 Calculation of income and life cycle revenue stream for cattle

Net income from cattle production was calculated using a net present value approach. We use this approach rather than a current snapshot based on cattle sales because a majority of cattle farms sold only 10% of their heard during 2010 and most herds were comprised largely of calves and juveniles, rather than adult cattle. The most common sale age for cattle in the region is 36 months, so most of the farms with high proportions of calves and juveniles and would not be selling their cattle for another one to two years. Our equation for this calculation is as follows:

$$\boldsymbol{NPV} = \mathbf{P_{2010,i}} \times \boldsymbol{S_{2010,i}} - \boldsymbol{C_{2010,i}} + \frac{\mathbf{P_{2011}} \times \boldsymbol{S_{2011}} - \boldsymbol{C_{2010}}}{\mathbf{1} + \boldsymbol{r}} + \frac{\mathbf{P_{2012}} \times \boldsymbol{S_{2012}} - \boldsymbol{C_{2010}}}{(\mathbf{1} + \boldsymbol{r})^{\mathbf{2}}}$$

where r is the interest rate, assumed to be 5%, $P_{year,i}$ is the cattle sale price in 2010, 2011, and 2012 received by each property-household, $S_{2010,i}$ is the reported number of cattle sold in 2009, $S_{2011,i}$ is the number of male and female juveniles reported on the property, and $S_{2012,i}$ is the number of male and female calves reported on the property. This calculation assumes that juveniles need on average 12 more months to mature before being sold and calves need 24 more months. It also assumes that all males and females are sold, rather than being kept for breeding. Total operating costs for the entire herd are assumed to be static across all three years (C_{2010}) because we lack any method to refine cost estimates in future periods. The price for each property-household in 2010 was obtained directly from the survey. Average local prices in 2010, 2011, and 2012 were obtained from local experts. We estimated prices received by each property-household in 2011 and 2012, by adding the % increase or decrease in the average local market price for cattle in 2011 and 2012 to the prices received by each farm in 2010. Off-farm income includes remittances, wage labor, and conditional cash transfers. We do not deduct household expenditures.

A2.2 Determinants of land use

To understand how assets influence land use choices among our households we use a multinomial logit model (Wooldridge, 2010). We assume that farmers choose their land use strategies seeking to maximize utility (subjective wellbeing) from available alternatives. Net benefits, *B*, for the farmer *i* using the alternative *j* (i.e., one of the five land uses) are assumed to be a deterministic function of the set of conditioning assets *Z* owned by the farmer (Table S1), so $B_{ij}(Z)$.

The expected utility *U* of a given alternative *j* is a function of the benefits of such assets, $U\left(B_{ij}(Z)\right)$, and can be decomposed into deterministic and stochastic elements, given by:

$$U\left(\mathrm{B_{ij}}(\mathrm{Z})\right) = f_{ij}(\mathrm{Z}) + \varepsilon_{\mathrm{ij}} \; \forall\, i=1,\cdots, N \; ; \; \forall\, j=1,\cdots, J, \quad \mathbf{[1]}$$

where $f_{ij}(Z)$ is a deterministic function of farm assets, ε_{ij} is a random variable representing unobserved farm attributes.

It is assumed that the farmer will continue to use alternative *j*, so long as the expected utility from *j* is at least as high as another alternative *k*, such that: $U(B_{ij}(Z)) \geq U(B_{ik}(Z)) \geq 0$. Thus, among a sample of *N* property-households, choosing among *j* mutually exclusive land use strategies, the probability that a property-household i selects alternative j (over all other alternatives), conditioned to Z, can be represented as:

$$Prob(\mathrm{y = j}) = Prob\Big(U\big(\mathrm{B_{ij}(Z), \varepsilon_{ij}}\big) > U(\mathrm{B_{i0}(Z), \varepsilon_{i0}}) >, \cdots, > U\big(\mathrm{B_{iJ-1}(Z), \varepsilon_{ij}}\big) > U\big(\mathrm{B_{iJ}(Z), \varepsilon_{ij}}\big)\Big) \quad \textbf{[2]}$$

Equation (2) can be rewritten as:

$$Prob(\mathrm{y = j}) = Prob\Big(U\big(\mathrm{B_{ij}(Z), \varepsilon_{ij}}\big) > U(\mathrm{B_{ik}(Z), \varepsilon_{ik}}), \forall\ \mathrm{k = 1, \ldots, J}\Big) \quad \textbf{[3]}$$

It is generally assumed (McFadden, 1980; Wooldridge, 2010) that the last probability is a non-linear function G(.) of only Z and the parameters of the linear approximation, β_j, j=1,…7, which can be subsumed to a J-column matrix, β, such that P(y = j|Z) = G(Z,β). The multinomial logit model corresponds to the following specification for G(Z,β):

$$G(\mathrm{Z}, \beta) = \begin{cases} \dfrac{\exp(\mathrm{Z}\beta_{\mathrm{j}})}{1+\sum_{\mathrm{k}=1}^{7}\exp(\mathrm{Z}\beta_{\mathrm{k}})}, \text{if j} = 2, \ldots, 7 \\ \dfrac{1}{1+\sum_{=1}^{7}\exp(\mathrm{Z}\beta_{\mathrm{k}})}, \text{if j} = 1 \end{cases} \quad \textbf{[4]}$$

So, $\sum_{j=1}^{5} Prob(y = j|Z) = 1$.

In operationalizing this framework we also include other household attributes (C) and a regional fixed effect (F) as control variables. All variables included in the model vary by individual (*i*), not by alternative (*j*).

The estimated model is:

$$y_{ij} = \alpha + \sum_{j=1}^{7}\beta_j Z_i + \sum_{j=1}^{7}\lambda_j C_i + \gamma F + \varepsilon_{ij} \quad \textbf{[5]}$$

where $y_{ij}$is the log[prob *(i* devoted to *j*)/prob (*i* devoted to the base alternative)]. That is to say that the coefficients of one land use strategy – in this case, “cattle” as the land use strategy – are normalized to zero. Thus, β and γ represent the effects on the log-odds between the alternative *j* and the base alternative. Finally, estimation of equation 5 is best carried out by maximum likelihood (Wooldridge, 2010).

In choosing a multinomial logit approach over a binary choice model we considered the following issues. The binary choice model implies that the unordered response has only two outcomes (e.g., Y = 1 for cattle and Y = 0 for all other things). If there are other possible different alternatives alongside of the two retained, this model assumes that when an individual has to make the choice, he takes his decision without considering any other possible forms of land use, or, that is to say, that he considers that all other alternatives may be gathered into Y = 0. This can be challenged because the choice of Y = 1 (“cattle”) can be different regarding the nature of the other possible alternative. For instance, the choice of cattle is obviously different if the farmer has to choose between cattle and soy or between cattle and specialty crops. Therefore, the multinomial logit model is preferred to the binary choice model because several different alternatives are naturally at stake. In using the multinomial logit the independence of irrelevant alternatives assumption, i.e. the relative probabilities for any two alternatives depend only on the attributes of those two alternatives, may be questioned. If this assumption is not valid, the choice of the multinomial logit model should be challenged and a hierarchical model (e.g. a nested logit model) or a conditional probit model should be preferred. However, the multinomial logit model used in this paper is not sensitive to the IIA hypothesis because explanatory variables differ only across individuals and not across alternatives (Wooldridge, 2010 – p. 501-504). Finally, the multinomial logit model has been used in several prior studies to investigate the determinants of land use in the literature (See Chomitz and Gray, 1996; Li et al., 2013; Schuck et al., 2002 as examples).

3. Determinants of subjective wellbeing

To understand how assets, land use, and income influence subjective wellbeing (represented here as an ordinal factor variable – "life quality") we utilize an ordered logistic model (McKelvey and Zavoina, 1975):

$$y_i^* = \alpha + \delta I_i + \eta LU_i + \beta Z_i + \lambda C_i + \gamma F + \varepsilon_i, \textbf{[6]}$$

where y_i^* is a latent (unobserved) measure of the subjective wellbeing of a given property-household *i*, LU_i is a set of categorical variables indicating the land use each household-property *i* pursues in terms of land uses, I_i represents total farm income and off-farm income as separate variables, Z_i, C_i, and *F* are the same set of variables as above, and ε_i is the stochastic error term.

The relationship between observed life quality, Y, and the latent variable behind the reported levels (the true continuous measure of subjective wellbeing), Y*, is as follows (Wooldridge, 2010):

If $Y^* \leq \delta_1$, then y = 1
If $\delta_1 < Y^* \leq \delta_2$, then y = 2
If $Y^* > \delta_2$, then y = 3

With δ_1 and δ_2 being the cut points on the latent variable behind the choice of life quality levels. Thus the probabilities for each of the observed ordinal responses (1, 2, 3) for low life quality, moderate life quality, and high life quality, respectively, will be given as:

$P(Y=1) = P(Y^* \leq \delta_1) = P(\boldsymbol{\beta}'x + \boldsymbol{\varepsilon}^* \leq \delta_1) = F(\delta_1 - \boldsymbol{\beta}'x)$
$P(Y=2) = P(\delta_1 < Y^* \leq \delta_2) = F(\boldsymbol{\delta_2} - \boldsymbol{\beta}'x) - F(\delta_1 - \boldsymbol{\beta}'x)$
$P(Y=3) = 1 - F(\boldsymbol{\delta_2} - \boldsymbol{\beta}'x)$

where *F* is the cumulative distribution function (CDF) for the stochastic error term ε, assumed to follow a logistic function. The unobservable cut points (δ_1 and δ_2) are estimated together with other parameters in the model. If an intercept is appended to the model, it automatically plays the role of the cut point. This model assumes that the relationship between each pair of life quality outcome groups (e.g. 1 to 2 and 2 to 3) is the same.