Pre-Analysis Plan: Trading off Investment in Human Capital versus Productive Assets: Perceptions of Risk and Returns

Yuma Noritomo* Karlijn Morsink† Kelvin Shikuku ‡

March 2026

1 Summary

This pre-analysis plan describes the design and analysis of a study that examines whether, and to what extent, information about climate risk and returns to education affects beliefs and investment decisions about livestock production and human capital accumulation in pastoralist communities in Kenya and Ethiopia. The study measures changes in beliefs and planned investment among approximately 2,000 households with school-aged boys, drawn from the sample of households surveyed in the study described in Morsink et al. (2025).

Following the guidance of Banerjee et al. (2020), we parsimoniously outline the study's outcomes, research questions, and empirical approach.

2 Experimental Design

2.1 Randomization

Randomization is conducted at the household level. Assignment to treatment and control groups is stratified by districts and unit areas of insurance[1] for the larger study, yielding 14 strata. Households are assigned to one of three treatment arms or to a control arm corresponding to the information interventions, with an expected sample size of approximately 500 households per group.

2.2 Interventions

This trial evaluates two information interventions implemented within the survey. During the survey, we first collect prior belief data, then provide randomly assigned information, and subsequently re-elicit beliefs for the treatment groups and measure outcomes. One intervention provides

*Charles H. Dyson School of Applied Economics and Management at Cornell University. Email: yn266@cornell.edu

†Utrecht University School of Economics, Development Economics Group at Wageningen University, Charles H. Dyson School of Applied Economics and Management at Cornell University, and Center for the Economics Analysis of Risk at Georgia State University.

‡International Livestock Research Institute.

[1] A unit area of insurance is the spatial unit at which premium rates are set, policies are sold, the drought index values are calculated, and for which payouts are determined.

information on climate risk, while the other provides information on labor market returns to education. We randomize households to receive either no information (control), climate risk information only, labor market information only, or both climate risk and labor market information jointly.

2.2.1 Climate Risk Information Intervention

We construct historical climate risk information using publicly available measures of the Normalized Difference Vegetation Index (NDVI), which has been shown to be highly correlated with livestock mortality in this setting (Chantarat et al., 2013).

Droughts are defined using NDVI data from 2000 to 2025. For each location and season (long rains/long dry season and short rains/short dry season), we calculate a standardized NDVI z-score relative to the historical distribution for a given location and a season. A season is classified as a drought if the z-score falls below the 20th percentile of the historical distribution. We then aggregate the number of droughts across seasons and years.

The information provided to respondents reports the number of droughts that occurred in their region during five-year intervals over the past 25 years: 2000–2004, 2005–2009, 2010–2014, 2015–2019, and 2020–2024. Respondents are also shown whether drought frequency appears to be increasing or decreasing by comparing the earlier period (2000–2009) with the more recent period (2015–2024).

2.2.2 Labor Market Information Intervention

To inform beliefs about returns to education, we use nationally representative data to characterize labor market conditions. In Kenya, we rely on the Kenya Continuous Household Survey (KCHS), using the most recent wave available from 2021.[2] In Ethiopia, we use the Socio-Economic Panel Survey 2021–2022.[3]

From these datasets, we construct measures of average wages for those who have employment, and employment probabilities for men aged 25-34, conditional on educational attainment. Employment is defined as having a primary income-earning activity as a paid employee outside the household in Kenya, and as employment in a wage job or own non-farm enterprise in Ethiopia. Earnings are defined as wages or gross salary from paid employment outside the household in Kenya, and as wages or salary payments in Ethiopia. By doing so we effectively deliver information about the expected earnings in the labor market conditional on different educational attainment.

3 Outcomes

We collect outcome data on (expected) investment in education and livestock within the same survey as the information interventions. Because longer-run outcomes are not immediately observable, we also elicit respondents' expected or planned investment decisions.[4] For continuous or count variables, outliers will be winsorized at the 99th percentile and 1 percentile from the bottom if necessary.

[2] Accessible at `https://statistics.knbs.or.ke/nada/index.php/catalog/123/related-materials`.

[3] Available at `https://microdata.worldbank.org/index.php/catalog/6161`.

[4] We plan to revise the PAP to include a follow-up data collection with the same households in order to assess longer-run realized outcomes.

3.1 Primary Outcomes

The primary outcomes are as follows:

- **Voucher choice.** Respondents are asked to choose between a voucher worth 2,000 KSH in Kenya or 1,000 ETB in Ethiopia for educational expenses (such as books or uniforms) and a voucher of the same value for livestock inputs (such as veterinary supplies); the outcome equals one if the education voucher is chosen and zero otherwise.
- **Future education and livelihood.** We elicit respondents' expectations about the future educational attainment (in years), expected occupation (labor for others, skill-based job, work in a shop, or white-collar employment in a private company, government, or NGO), expected earnings (reported in bins of 0–4,000; 4,001–7,000; 7,001–12,000; 12,001–17,000; and 17,001+ ETB in Ethiopia, or 0–14,999; 15,000–30,499; 30,500–56,499; 56,500–97,499; and 97,500+ KSH in Kenya), and a binary indicator of whether the child is expected to migrate to secondary towns for work in 20 years, for one randomly chosen school-aged boy in the household.
- **Planned child activities.** We elicit expected child activities (herding household-owned livestock; livestock production, e.g., milking and sale of livestock products; livestock trading/brokerage; petty trading, e.g., charcoal or water trading; shop/business owner; unpaid work in the household's shop/business; casual labor, e.g., herding for pay or farming for pay; wage/salaried employment at a government institution; wage/salaried employment in the private sector; wage/salaried employment at an NGO; farming non-livestock crops; house/domestic work; student; not working: too old; not working: too young; not working: unable; looking for a job; fishing; poultry production; mining; motorbike rider/operator) for the same randomly selected school-aged boy in upcoming seasons, both when schools are in session and during school holidays.

3.2 Secondary Outcomes

Secondary outcomes capture beliefs and expected livestock investment.

- **Beliefs about climate risk.** We measure historical and future beliefs about climate risk. In particular, we measure respondents' reported number of droughts experienced in their location during the past 25 years and whether respondents believe the pattern of droughts has increased or decreased. We also measure the expected number of droughts and the expected pattern of droughts over the next 25 years.
- **Beliefs about returns to education.** Beliefs about returns to education are measured using respondents' perceived likelihood of securing employment, reported on a 0–10 scale, and expected earnings conditional on employment, reported in the same bins described above, for skill-based jobs (e.g., carpenter, mechanic, electrician, driver, beekeeping, construction/masonry, hairdressing, tailoring, handicrafts/beadwork, leathermaking), intermediary jobs (e.g., shopkeeper, untrained teacher, office assistant), and white-collar jobs (e.g., teacher, NGO jobs, government jobs, private sector jobs), conditional on different levels of educational attainment (primary, secondary, and tertiary) and location (their community or a relevant secondary town). From these measures, we construct expected earnings as the product of the perceived employment probability and expected earnings, and then take the difference in expected earnings between higher and lower levels of educational attainment for a given job type and location.

- **Expected livestock investment.** We measure expected expenditures on livestock-related inputs in the next year. These investments are recorded by input type (water for animals; animal feed/fodder (excluding concentrate, pellets, and minerals); vaccinations for animals; tick and fly control for livestock; other veterinary services; transportation/other transaction costs for animals; livestock tax; cash payments to herdsmen/boys (excluding household members); in-kind payments to herdsmen (excluding household members); fees to access pasture land) and season (short dry, long rain, long dry, and short rain seasons), and are aggregated into annual monetary values.

4 Analysis

We examine the following research questions.

- Research question 1 ("first stage"): To what extent does the provision of information about climate risks and labor market conditions lead respondents to update their beliefs about current and future returns to education and livestock production?
- Research question 2 (ITT or "second stage"): To what extent does this information and beliefs affect expected investment in children's education or child labor and expected livestock investment?

4.1 Effects on beliefs

We will estimate models of the following form:

$$\begin{aligned} y_{is} =& \beta_c T_i^c (x^c - x^c_{i,\text{prior}}) + \beta_l T_i^l (x^l - x^l_{i,\text{prior}}) + \beta_{b,c} T_i^b (x^c - x^c_{i,\text{prior}}) + \beta_{b,l} T_i^b (x^l - x^l_{i,\text{prior}}) \\ &+ \sum_{k \in \{c,l,b\}} \gamma_k T_i^k + \sum_{k \in \{c,l\}} \zeta_k (x^k - x^k_{i,\text{prior}}) + \delta_s + \varepsilon_{is} \end{aligned} \tag{1}$$

where y_{is} denotes posterior beliefs for household i in stratum s, measured in the survey for the control group and after the information intervention within the same survey for the treatment group. One of three possible information signals is randomly provided to respondents: the treatment indicators T_i^c, T_i^l, and T_i^b are dummy variables indicating that household i receives each signal–climate risk information c, labor market information l, or both b, respectively–and zero otherwise. Households in the control group do not receive any information. x^c and x^l are the information signals received for climate risk and labor market, while $x^c_{i,\text{prior}}$ and $x^l_{i,\text{prior}}$ are the prior beliefs in the same unit of measure. The difference between the signal and the respondent's prior about the signal captures the perception gap. δ_s captures randomization stratum fixed effects. We will include household-level control variables if there is imbalance in the randomization. ε_{is} is an idiosyncratic error term.[5,6] Since randomization occurs at the household level, we compute robust standard errors following Abadie et al. (2022). Given the rapid evolution of methodology in this space, throughout the analysis we will use frontier methods to address multiple hypothesis testing including the use of indices where relevant. We also test whether $\beta_k = \beta_{b,k}$ for $k \in \{c,l\}$.

[5]Since heterogeneous belief updating can attenuate average treatment effects when individuals revise their beliefs in opposite directions, we estimate the intention-to-treat (ITT) specification only in cases where the majority of baseline beliefs lie on the same side of the information signal (Haaland, Roth, and Wohlfart, 2023).

[6]We explore the higher order interaction effects if we see the coefficients suggest such effects.

4.2 Effects on investment

Following Fuster and Zafar (2023), we leverage the exogenous variation from the randomized information treatments, which shift respondents' expectations, to study the effects on subsequent economic behavior. We estimate the following models:

$$\begin{aligned} a_{is} =& \beta_c T_i^c (x^c - x^c_{i,\text{prior}}) + \beta_l T_i^l (x^l - x^l_{i,\text{prior}}) + \beta_{b,c} T_i^b (x^c - x^c_{i,\text{prior}}) + \beta_{b,l} T_i^b (x^l - x^l_{i,\text{prior}}) \\ &+ \sum_{k \in \{c,l,b\}} \gamma_k T_i^k + \sum_{k \in \{c,l\}} \zeta_k (x^k - x^k_{i,\text{prior}}) + \delta_s + \varepsilon_{is} \end{aligned} \tag{2}$$

where a_{is} denotes the investment outcomes of interest: voucher choice, planned education and livelihood, child activities, and livestock investment. We control for baseline outcomes whenever available.[7] If the first stage is strong enough for conventional threshold in equation 1 or the corresponding ITT specification, we also estimate the effects of beliefs on actions with the following second stage:

$$a_{is} = \kappa_c \widehat{y}^c_{is} + \kappa_l \widehat{y}^l_{is} + \sum_{k \in \{c,l,b\}} \gamma_k T_i^k + \sum_{k \in \{c,l\}} \zeta_k (x^k - x^k_{i,\text{prior}}) + \delta_s + \varepsilon_{is} \tag{3}$$

where the posterior beliefs $\widehat{y}^c_{is}$ and $\widehat{y}^l_{is}$ are instrumented by information treatments in equation 1 or similar ITT specification. For instrument validity, we will test for monotonicity and run appropriate regressions with heterogeneity by prior beliefs relative to the signal or by subsamples.

4.3 Heterogeneity

We will conduct analyses examining heterogeneous treatment effects by the baseline variables listed below.

- Prior beliefs relative to the signal
- Livestock size and composition
- Household composition
- Household has any non-farm employment or income
- Education and gender of household head

4.4 Exploratory analysis

We plan to conduct two exploratory analyses. First, we explore the possibility that the effects of belief updates may be nonlinear and that there are interaction effects when both pieces of information are provided. Second, we decompose the effects of expected wages by examining how wage information and employment probability information affect belief updating and outcomes separately.

[7] Similar to the discussion above, we estimate the ITT specification when baseline beliefs relative to the information signal are concentrated on one side.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge (2022). "When Should You Adjust Standard Errors for Clustering?" *Quarterly Journal of Economics* 138.1, pp. 1–35.

Banerjee, Abhijit, Esther Duflo, Amy Finkelstein, Lawrence F Katz, Benjamin A Olken, and Anja Sautmann (2020). "In Praise of Moderation: Suggestions for the Scope and Use of Pre-Analysis Plans for RCTs in Economics". *NBER Working Paper* No. w26993.

Chantarat, Sommarat, Andrew G. Mude, Christopher B. Barrett, and Michael R. Carter (2013). "Designing Index-Based Livestock Insurance for Managing Asset Risk in Northern Kenya". *Journal of Risk and Insurance* 80.1, pp. 205–237.

Fuster, Andreas and Basit Zafar (2023). "Survey experiments on economic expectations". In: *Handbook of Economic Expectations*. Ed. by Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw. Academic Press, pp. 107–130.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart (2023). "Designing Information Provision Experiments". *Journal of Economic Literature* 61.1, pp. 3–40.

Morsink, Karlijn, Lotte Van Der Haar, Nathan Jensen, Kelvin Shikuku, Hyuk Son, and Tagel Gebrehiwot (2025). *The Impact of Information about Consumer Value and Contract Performance on Demand and Welfare of Insurance*. Pre-Analysis Plan.