

Causal Effects of Climate-Smart Agriculture Adoption on Smallholder Welfare and Food Security in Zambia: Evidence from Endogenous Switching Regression, Double Machine Learning, and Bayesian Estimation

Evans Kudzai Chinembiri

ewkchinembiri@gmail.com

North-West University <https://orcid.org/0009-0003-7304-0214>

Munyaradzi Mutenje

International Water Management Institute <https://orcid.org/0000-0002-7829-9300>

Nkele D. Malahlela

North-West University <https://orcid.org/0009-0001-1653-7553>

Research Article

Keywords: climate-smart agriculture, food security, double machine learning, endogenous switching regression, smallholder farmers, Zambia

Posted Date: September 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10994635/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare potential competing interests as follows: M. Mutenje is employed by the International Water Management Institute (IWMI), which owns the household survey data analysed in this study and received a United States Agency for International Development grant that funded its collection. E.K. Chinembiri was previously employed by IWMI and is affiliated with North-West University, where this analysis was conducted as doctoral research under ethics clearance NWU-00551-25-A9. Beyond M. Mutenje's co-authorship contribution, which includes data collection as recorded in the authors' contributions, IWMI had no role in designing the analysis, interpreting results, or deciding to publish, and did not review or approve the manuscript. The authors declare no financial interest in the outcomes reported.

Abstract

Climate-smart agriculture (CSA) is promoted across sub-Saharan Africa to raise smallholder productivity, resilience, and welfare, with food security the ultimate objective, yet rigorous causal evidence for Zambia remains limited. Using 2,116 smallholder households from the Water and Soil Accelerator (WASA) survey across four provinces, this study estimates the causal effects of CSA adoption on agricultural income and crop yield, treating both as proximate determinants of household food access and availability. Three complementary estimators are applied. Multinomial Endogenous Switching Regression (MESR) models selection into four ordered adoption regimes (T0 to T3) and yields stratum-specific effects on the treated (ATTs). Double Machine Learning (DML) estimates population-average effects through Neyman-orthogonal, cross-fitted regression with LASSO and Random Forest learners. A Bayesian DML extension corroborates these, with the posterior probability of a positive effect at 1.000 for all outcomes and posterior and frequentist means differing by at most 0.041. A Stable Unit Treatment Value Assumption (SUTVA) test confirms spillover below the 10 percent materiality threshold. DML places the income premium at approximately 33 percent and the yield premium at approximately 47 percent for the smallholder population, both significant at the 1 percent level. MESR traces a consistent positive gradient, proportionally larger for subsistence than for commercial farmers and falling at the highest adoption intensity; after a full two-stage bootstrap, only the subsistence yield gains at dual- and multi-practice intensity reach individual significance, so identification rests on DML and BDML. The results support targeted CSA promotion for subsistence households and caution against universal multi-practice mandates.

1. Introduction

Agriculture remains the primary livelihood source for the majority of rural households across sub-Saharan Africa, accounting for the bulk of employment, food security, and household income for smallholder farmers who face compounding pressures from climate variability, input cost inflation, and constrained market access. In Zambia, approximately 70 percent of the population depends on rain-fed smallholder agriculture, yet agricultural productivity has remained persistently below potential (Ng'ombe et al., 2017). Climate-smart agriculture (CSA), a package of practices including conservation agriculture, agroforestry, residue management, integrated soil fertility management, and assisted natural regeneration, has emerged as a central policy response, actively promoted by the Zambian government, development finance institutions, and international agricultural research systems.

The welfare that these programmes ultimately target is food security, not income or yield in isolation. Food security is a multidimensional condition spanning availability, access, utilisation, and stability, and household income and own-farm productivity are two of its most direct upstream determinants for rural producers (FAO, 2009). For smallholder households that both produce and purchase food, a gain in crop income raises the budget available to acquire a more diverse diet, while a gain in own-farm yield raises the quantity of staples retained for home consumption; both channels have been shown to translate into measurable improvements in household food access and dietary diversity in sub-Saharan settings

(Kassie et al., 2015). This study measures the two upstream determinants directly and treats them as the food-security pathway through which CSA adoption operates, while remaining explicit that neither is itself a consumption measure.

The Water and Soil Accelerator instrument used here did capture a household food-consumption module, comprising meals-taken-yesterday counts for adult men, adult women, and children, and a multi-item food-frequency recall by food group and source. These items support a future household dietary-diversity or food-consumption-score analysis. The present study does not estimate dietary outcomes; it analyses agricultural income and crop yield as the welfare outcomes for which a complete causal-identification design could be constructed, and it justifies them as validated proxies for the food-access and food-availability dimensions rather than as substitutes for direct consumption measurement (Carletto et al., 2013). A companion analysis of the dietary module is identified as the natural next step.

Despite substantial investment in CSA promotion and extension, rigorous causal evidence on welfare effects in Zambia stays sparse. The existing literature is dominated by propensity score matching and OLS-based approaches that assume selection on observables and cannot fully address unobserved heterogeneity in adoption decisions. Previous studies have rarely engaged with the ordered, multi-practice nature of CSA uptake. Treating adoption as a binary indicator collapses the heterogeneous welfare pathways available to different types of smallholder households.

Three methodological contributions to the CSA impact literature follow. First, the analysis applies Multinomial Endogenous Switching Regression (MESR) to model selection into four ordered adoption regimes, non-adoption (T0), single-practice adoption (T1), dual-practice adoption (T2), and multi-practice adoption (T3), and estimates adoption-intensity-specific average treatment effects on the treated (ATTs), disaggregated by farm stratum (subsistence versus commercial). Second, it applies Double Machine Learning (DML) to estimate population-average treatment effects through Neyman-orthogonal, cross-fitted semi-parametric regression, providing a methodological counterpoint to the parametric MESR framework. Third, a Bayesian DML estimator provides full posterior distributions over treatment effect parameters, with convergence diagnostics confirming satisfactory sampling ($\hat{R}$ at most 1.0015; effective sample size exceeding 4,900 per parameter).

Triangulation across methodologically distinct estimators, one parametric and structural, one semi-parametric and machine-learning-based, and one fully Bayesian, constitutes the evidentiary strategy of this study. The data are drawn from the WASA Project household survey, comprising 2,116 smallholder farming households across four provinces of Zambia. Outcomes of interest are agricultural income (total farm household income from crop production, in Zambian Kwacha) and crop yield (ZMW per hectare).

The remainder of the paper is organised as follows. Section 2 reviews the literature on CSA adoption and impact evaluation methodology. Section 3 presents the conceptual framework. Section 4 describes the data and study area. Section 5 sets out the estimation strategies. Section 6 reports the results. Section 7 discusses the findings in relation to the existing literature and policy implications. Section 8 concludes.

2. Literature Review

The theoretical case for CSA rests on simultaneous productivity and resilience gains from practices that reduce soil degradation, improve water retention, and enhance nutrient cycling. A growing body of empirical evidence documents positive yield effects in experimental and quasi-experimental settings across sub-Saharan Africa (Arslan et al., 2014; Ng'ombe et al., 2017), but farm-level adoption rates stay stubbornly below programme targets. The gap between demonstrated effectiveness and realised adoption has motivated parallel literatures on both adoption determinants and causal impact evaluation.

On adoption determinants, Nhemachena and Hassan (2007) demonstrated that access to credit, extension services, and market information are significant predictors of CSA uptake in Southern Africa. Asfaw et al. (2012) found that adoption of modern agricultural technologies in Ethiopia and Tanzania was associated with significant income gains but varied substantially by household resource endowment, with wealthier households better positioned to capture returns. These findings established the conceptual importance of heterogeneous adoption returns, a dimension that binary impact evaluation frameworks systematically obscure.

The food-security stakes of this adoption gap are well established for the region. CSA and sustainable-intensification practices have been linked not only to higher farm income but to improved household food security and dietary diversity in several sub-Saharan settings, with the income channel and the own-production channel both contributing (Lipper et al., 2014; Ogisi and Begho, 2023). Zambia-specific evidence on rural food insecurity and seasonal food gaps motivates a welfare evaluation that keeps food access in view even where the analysed outcomes are income and yield (Sitko, 2007; Kitsuki and Sakurai, 2023).

On impact evaluation methodology, Teklewold et al. (2013) addressed heterogeneous adoption in Ethiopia using a multinomial switching regression framework, finding positive and significant treatment effects on both income and food security, with adoption-intensity gradients consistent with increasing, though diminishing, returns to multi-practice engagement. Manda et al. (2017) estimated ex-ante welfare impacts of maize-soybean rotation adoption in eastern Zambia, projecting positive income effects for adopting households.

The application of machine learning methods to causal inference in agricultural economics is more recent. Double machine learning, formalised by Chernozhukov et al. (2018), addresses the limitations of parametric structural models by using flexible, data-adaptive learners, Least Absolute Shrinkage and Selection Operator (LASSO), Random Forest, and gradient boosting, to estimate nuisance functions for the outcome and treatment propensity, then recovering the structural treatment effect parameter through Neyman-orthogonal moment conditions and cross-fitting. Bach et al. (2024) implemented the DoubleML framework in R, facilitating its application in development economics contexts. Applications to agricultural welfare evaluation in sub-Saharan Africa stay limited, creating a methodological gap that this study addresses directly.

Bayesian approaches to treatment effect estimation offer a further methodological frontier. DiTraglia and Liu (2025) extended the DML framework to a fully Bayesian setting, providing posterior distributions over treatment effects rather than point estimates with frequentist confidence intervals. This is particularly valuable for policy-oriented inference, where decision-makers must weigh not only the magnitude but also the uncertainty of estimated returns. To the authors' knowledge, the DiTraglia and Liu Bayesian DML estimator has not previously been applied to CSA adoption contexts, making this study one of the first such applications in the sub-Saharan African agricultural economics literature, while recognising that the estimator itself rests on a recent and as-yet-unpublished methodological proposal.

Beyond methodology, two empirical findings from the existing literature frame this study's interpretive approach. The income and yield effects of CSA adoption appear heterogeneous across farm typologies, with subsistence-oriented households consistently showing larger proportional gains than commercially oriented counterparts (Teklewold et al., 2013). Community-level spillover from CSA adoption, through information diffusion, input sharing, and social network demonstration effects, can confound household-level treatment effect estimates, motivating formal SUTVA testing (Miguel and Kremer, 2004). This study addresses both dimensions directly.

3. Conceptual Framework

The study is anchored in the Sustainable Livelihoods Framework (SLF; Chambers and Conway, 1992), which conceptualises agricultural decision-making as a function of household capital endowments across five dimensions: human capital (education, health, skills), natural capital (land, water, agro-ecological conditions), financial capital (savings, credit access, remittances), physical capital (equipment, irrigation infrastructure, storage), and social capital (networks, group membership, extension access). Within this framework, CSA adoption is modelled as an ordered choice along an adoption intensity continuum, shaped by the household's capital portfolio and mediated by institutional context, particularly the agricultural camp system through which extension is organised in Zambia.

Figure 1 renders the conceptual model as a circular, four-stage process. Capital Endowment (stage 1) feeds Awareness and Evaluation (stage 2), which leads to the Adoption Decision (stage 3), bifurcated into single-practice and bundled multi-practice adoption, and on to Welfare and Resilience Outcomes (stage 4). A feedback arrow returns from stage 4 to stage 1, capturing how realised welfare gains rebuild household capital and enable higher-intensity adoption in later seasons. Zambia-specific institutional mediators, the agricultural camp extension system, agro-ecological risk across the four study provinces, and community institutions such as NGOs and cooperatives, moderate the stage-two-to-stage-three pathway. No numeric values appear in the figure.

The welfare production function is specified as follows. Let Y_i denote household i 's welfare outcome, agricultural income or crop yield, and let $T_i \in \{T_0, T_1, T_2, T_3\}$ denote adoption intensity, where T_i records how many distinct CSA practice categories the household has adopted. The causal estimand of interest, the average treatment effect on the treated (ATT), is the expected welfare differential between a

household's realised outcome under its observed adoption status and the counterfactual outcome it would have achieved under the reference (T_0) status, conditional on its capital endowment vector X_i , the set of household, farm, financial, social, and institutional characteristics described in Section 4.

Two counterfactual estimands support inference. The Average Treatment Effect (ATE) measures the expected gain from moving a randomly selected household from non-adoption to adoption, relevant for universal programme accessibility questions. The Average Treatment Effect on the Treated (ATT) restricts inference to adopting households, more informative for ex-post programme evaluation and targeting. The DML framework estimates both; the MESR framework estimates the ATT by adoption regime.

The conceptual model predicts three empirical patterns. CSA adoption should generate positive welfare returns for most households, though the magnitude will vary by capital endowment, particularly land quality, extension access, and farm income diversification. Adoption-intensity returns may be non-monotonic: initial practice adoption may capture the largest welfare gains, with subsequent practices adding positive but diminishing returns. Subsistence and commercial farmers may show divergent yield responses, reflecting different production objectives and input allocation strategies.

The model synthesises and extends three established theoretical traditions. The Sustainable Livelihoods Framework (Chambers and Conway, 1992) provides the capital endowment architecture: human, natural, financial, physical, and social capitals jointly determine a household's structural capacity to act on adoption intentions. Rogers' (2003) diffusion of innovations tradition provides the stage-based sequential logic, in which adoption proceeds through linked phases of awareness, evaluation, decision, and confirmation. The agricultural technology adoption literature (Feder et al., 1985) establishes the role of farm size, credit access, risk aversion, and institutional support as structural determinants of adoption behaviour.

Three specific extensions of these foundations constitute the conceptual contribution. The Adoption Decision stage is explicitly bifurcated into single-practice and bundled adoption, a distinction not formalised in prior CSA adoption frameworks for sub-Saharan Africa. This bifurcation is the theoretical basis for the ordered intensity variable T_0 through T_3 , and it is what makes the declining ATT gradient empirically interpretable rather than a statistical artefact. The framework is circular rather than linear: positive welfare and resilience outcomes at stage four feed back into improved household capital at stage one, enabling higher-intensity adoption in subsequent seasons. This feedback architecture generates both the poverty trap at T_0 and the pathway out of it as a dynamic process rather than a static state. The framework also incorporates Zambia-specific institutional mediators absent from generic adoption models: the agricultural camp system as the primary extension delivery channel, agro-ecological risk heterogeneity across the four study provinces, and community institutions including NGOs and cooperatives as moderators of the capacity-to-decision pathway. Together, these three extensions provide the theoretical scaffolding within which the MESR and DML causal estimates are interpreted.

4. Data and Study Area

The analysis draws on household survey data collected under the Water and Soil Accelerator (WASA) Project in Zambia. The survey was administered across four provinces, Central, Eastern, Southern, and Copperbelt, covering 2,116 smallholder farming households across 16 districts, surveyed by 27 enumerators. Data collection ran through ODK Collect between 29 November and 13 December 2025. The sampling design targets households engaged in rain-fed smallholder agriculture within agricultural camp service areas, enabling analysis of both individual adoption decisions and camp-level adoption spillover. Figure 2 locates the four study provinces within Zambia.

Survey modules capture CSA practice adoption (conservation agriculture, agroforestry, residue management, integrated soil fertility management, assisted natural regeneration), household demographics, farm characteristics, income sources, institutional access, and household food consumption. The food-consumption module records meals taken in the previous day for adult men, adult women, and children, and a multi-item food-frequency recall by food group and source; these items are reserved for a companion dietary-diversity analysis and are not estimated here. The CSA adoption intensity variable is constructed as an ordered categorical indicator: T0 (zero practices), T1 (one practice), T2 (two practices), and T3 (three or more practices), consistent with the adoption typology used in Teklewold et al. (2013).

Two welfare outcome variables are analysed. Farm income captures total household income derived from crop production, expressed in Zambian Kwacha (ZMW). This constitutes the primary income outcome, as it reflects the direct agricultural welfare pathway through which CSA adoption operates. Crop yield (ZMW per hectare) captures productivity effects on the farm. Both outcomes are log-transformed prior to analysis to address right-skewed distributions, and outliers are addressed through 99.5th-percentile top-coding, a standard treatment for heavy-tailed agricultural income and productivity data. Because the outcomes enter the models in logs while the MESR treatment effects are reported in ZMW levels, the level estimates are recovered from the log-scale predictions using the Duan (1983) nonparametric smearing retransformation, which corrects the bias that naive exponentiation introduces at heavy-tailed baselines by scaling each arm's exponentiated prediction by the mean of its exponentiated residuals. Statistical significance is set on the log-scale second stage and is unaffected by the retransformation; the smearing correction refines the reported Kwacha-level ATTs and the implied proportional gains, most materially at the low subsistence baselines where naive exponentiation is least reliable. Section 6 reports the Duan-smear level effects throughout.

The analytic sample for the MESR models is $N = 2,055$: the 2,116 surveyed households less the Vulnerable Cohort, which is held out of the switching-regression frame. The DML analytic sample is $N = 2,016$, drawn from the full surveyed file by listwise deletion of 100 observations with missing camp-level adoption data required for the SUTVA sensitivity test. That reduction is roughly 4.7 percent, unrelated to the outcome variables, which supports a missing-at-random treatment for the DML estimates. Farm

typology follows a principal-component-based subsistence-commercial index, with adopting households classified as subsistence-oriented (n = 911) or commercially oriented (n = 411).

Table 0

reports aggregated descriptive statistics for the MESR analytic sample (N = 2,055; Vulnerable Cohort excluded). Means appear with standard deviations in parentheses, and adoption-regime figures are shares.

Variable	Mean or share	Units or note
Agricultural income (crop)	40,418 (63,559)	ZMW
Crop yield	6,790 (7,670)	ZMW per hectare
Adoption regime T0 (non-adoption)	35.7	percent of households
Adoption regime T1 (single practice)	24.7	percent of households
Adoption regime T2 (dual practice)	22.0	percent of households
Adoption regime T3 (multi practice)	17.7	percent of households
Subsistence-oriented adopters	911	households
Commercially-oriented adopters	411	households
Household head age	48.8 (14.4)	years
Household head education	7.4 (3.4)	years of schooling
Farm size	8.0 (19.0)	hectares
<i>Note: Aggregated summary statistics only; no household-level or identifying information is reported. N = 2,055 (MESR analytic frame, Vulnerable Cohort excluded). Means are shown with standard deviations in parentheses; adoption-regime figures are shares of households. An extension-contact row is omitted: the underlying variable is a count of training and extension engagements whose minimum is one in this sample, so a "share with contact" of 100 percent carries no descriptive information. A distance-to-agricultural-camp row is omitted because the WASA dataset contains no such variable. Source: Authors' own calculations from WASA Project Survey data.</i>		

5. Methodology

This study estimates a single causal quantity, the welfare gain attributable to CSA adoption, through three estimators that rest on different assumptions and would fail in different ways. Reporting all three is the evidentiary point: where a structural selection model, a machine-learning estimator, and a Bayesian estimator agree, the conclusion does not hinge on the assumptions peculiar to any one of them. Technical Box 1 states the intuition behind all three in plain terms; the subsections then give the formal detail.

Technical Box 1: Three estimators in plain terms

MESR (Multinomial Endogenous Switching Regression) compares each adopter to a statistically reconstructed version of the same household had it not adopted, while correcting for the fact that adopters differ from non-adopters in ways the analyst cannot observe. It first models which adoption regime a household chooses, then uses that model to remove the part of the welfare gap that is due to who adopts rather than to adoption itself.

DML (Double Machine Learning) lets flexible machine-learning models absorb the many confounding characteristics of a household, so the adoption effect is isolated without the analyst having to guess the correct functional form. The "double" refers to modelling both the outcome and the probability of adopting, then combining them so that small errors in either one do not bias the final effect.

BDML (Bayesian DML) restates the same estimate as a full probability distribution rather than a single number with a confidence interval, so the analyst can read off directly how probable it is that the effect is positive and how concentrated that probability is.

The three estimators answer slightly different questions, MESR by adoption intensity and farm type, DML for the population and for adopters, BDML as a probability statement, and their agreement is what licenses a causal reading.

5.1 Multinomial Endogenous Switching Regression

In intuition, MESR asks a counterfactual question that a simple comparison of adopters and non-adopters cannot answer: what would an adopting household have earned had it not adopted, given that the households which choose to adopt are not a random sample? It answers this in two stages, modelling the adoption choice first and then correcting the outcome equations for the selection revealed by that choice.

The MESR framework (Bourguignon et al., 2007; Teklewold et al., 2013) consists of two stages. In the first stage, a multinomial logit model estimates household selection into the four adoption regimes (T0, T1, T2, T3) as a function of the full SLF covariate matrix X_i . One variable is held out of the second-stage outcome equations as the candidate exclusion restriction: the camp-level leave-one-out mean CSA adoption rate, which captures the adoption environment surrounding a household without entering its own outcome mechanically. Extension contact enters both stages as a control rather than as an excluded instrument, and the dataset carries no measure of distance to the agricultural camp. The first-stage estimates yield selection correction terms, the generalised inverse Mills ratios λ_j for each regime j , which summarise the correlation between the unobservables driving adoption and those driving the outcome.

Identification in this design does not rest on a clean exclusion restriction, and the paper is explicit about that. The natural candidate instrument, the camp-level adoption rate, fails a direct test of excludability. Added to the non-adopting (T0) outcome equation, where a valid excluded term should carry no direct effect, it is significantly associated with both farm income ($\beta = 1.17$, $p = 0.016$) and crop yield ($\beta = 2.61$, p

= 0.023), and a joint Wald test rejects exclusion for each outcome (income $\chi^2 = 5.78$, $p = 0.016$; yield $\chi^2 = 5.18$, $p = 0.023$). The reading is that the camp adoption rate is relevant but not excludable: it plausibly proxies the same community-level conditions that shape conditional returns. MESR identification therefore leans on the selection-correction functional form. The generalised inverse Mills ratios enter the outcome equations nonlinearly, so the model is identified by that nonlinearity even without an excluded instrument (Bourguignon et al., 2007). This is a weaker footing than an exclusion restriction would provide, and the stratum-specific ATTs are accordingly read alongside, and corroborated by, the Double Machine Learning and Bayesian estimators of Sections 5.2 and 5.3, which require no instrument and rest instead on conditional independence. On relevance, the excluded term is strongly associated with adoption intensity, and a joint likelihood-ratio test rejects irrelevance ($\chi^2 = 124.5$, $p < 0.001$). The first stage is quasi-separated on the camp adoption rate, so first-stage Wald inference is neither reported nor relied upon.¹

In the second stage, regime-specific outcome equations for farm income and crop yield are estimated by Ordinary Least Squares (OLS), augmented with the selection correction terms λ from the first stage. This correction accounts for correlation between unobservables that jointly determine adoption status and welfare outcomes, addressing the primary source of selection bias in cross-sectional impact evaluation. ATT estimates are computed as the expected welfare differential between observed and counterfactual outcomes for each adopting household, conditional on regime membership and farm typology.

The adoption intensity variable follows the CSA bundle taxonomy of Amadu et al. (2020), which classifies smallholder CSA practices into six theoretically grounded categories: (i) residue addition, manure application, mulching, and organic fertilisation; (ii) non-woody cultivation, intercropping, cover cropping, crop rotation, and legume integration; (iii) assisted regeneration, natural regeneration, cessation of burning, and minimum tillage; (iv) woody cultivation, tree planting and agroforestry; (v) physical infrastructure, irrigation, water harvesting structures, and soil conservation bunds; and (vi) mixed and precision measures, micro-dosing, integrated pest management, mechanisation, agricultural insurance, and weather information use. T0 denotes non-adoption of any category; T1 denotes adoption of exactly one category; T2 denotes adoption of any two categories; and T3 denotes adoption of three or more categories. The intensity variable therefore captures the breadth of CSA engagement rather than depth within a single category, consistent with the portfolio-based conceptualisation of CSA adoption in the literature (Teklewold et al., 2013; Arslan et al., 2014).

5.2 Double Machine Learning

In intuition, DML removes the influence of the many household characteristics that jointly drive adoption and welfare by letting machine-learning models predict each, then studying what is left. The estimator is built so that errors in those predictions do not transmit into bias for the adoption effect, which is what permits flexible learners without sacrificing valid inference.

Double Machine Learning (Chernozhukov et al., 2018) estimates causal treatment effect parameters while staying agnostic about the functional form of the relationship between high-dimensional covariates and the outcome or treatment propensity. Two model specifications are estimated. The Partially Linear Regression (PLR) model (Robinson, 1988) treats the binary adoption indicator as the treatment variable and estimates the population ATE through Neyman-orthogonal moment conditions, with LASSO and Random Forest nuisance learners for the outcome and propensity-score equations. The Interactive Regression Model (IRM) with the Augmented Inverse Probability Weighting (AIPW) score estimates the ATT, allowing for full treatment heterogeneity. The Neyman-orthogonality property means the moment condition used to estimate the adoption effect is locally insensitive to small errors in the machine-learning nuisance estimates; cross-fitting, in which nuisance functions are trained on one data partition and applied to another, removes the overfitting bias that would otherwise arise from using the same observations twice.

Both models use 5-fold cross-fitting (`set.seed(419)`) and ensemble nuisance learners to achieve root-n consistent estimation despite the use of regularised machine learning methods. The DoubleML R package (Bach et al., 2024) implements both specifications. Standard errors are obtained through the Neyman-orthogonal influence function, which is robust to first-stage estimation error, providing valid inference without sample-splitting adjustment.

The DML estimates rest on conditional independence (unconfoundedness): conditional on the covariate matrix, adoption is as good as randomly assigned. This assumption cannot be tested directly. Its plausibility here rests on the breadth of the SLF covariate matrix, which captures the human, natural, financial, physical, and social capital dimensions that the adoption literature identifies as the principal drivers of both selection and welfare. To bound the residual exposure to an omitted confounder, a sensitivity analysis quantifies how strong an unobserved confounder would have to be, in terms of its association with both adoption and the outcome, to overturn the estimated effect (Oster, 2019; Cinelli and Hazlett, 2020). That analysis was run on the linear specification carrying the DML control set. The robustness value is 0.152 for income and 0.146 for yield: an unobserved confounder would have to explain about 15 percent of the residual variation in both adoption and the outcome to move the point estimate to zero. The corresponding E-values are 1.90 and 2.04, and Oster's δ^* is 3.63 for income and 6.27 for yield, both far above the value of one at which the effect would look fragile. Selection on unobservables would need to run several times stronger than selection on the rich observed covariate set to nullify the effect, which places the estimates on robust, if not unassailable, ground.

5.3 Bayesian Double Machine Learning

In intuition, Bayesian DML produces the same adoption effect as the frequentist DML but expresses the answer as a probability distribution over possible effect sizes. That lets a policy reader ask directly how likely the effect is to be positive and how tightly the probability concentrates, rather than interpreting a confidence interval.

The Bayesian Double Machine Learning (BDML) estimator (DiTraglia and Liu, 2025) extends the frequentist DML framework to a fully Bayesian setting by placing prior distributions over the structural treatment effect parameters and recovering posterior distributions through Markov Chain Monte Carlo (MCMC). The specification implements the DiTraglia and Liu (2025) formulation in Stan through the brms interface. A weakly informative Normal(0, 1) prior is placed on the treatment effect for both outcomes, providing regularisation without constraining the posterior; the prior is centred at zero, so it does not push the estimate toward a positive result. The seed is fixed at 419 for reproducibility. Convergence is assessed through the Gelman-Rubin diagnostic ($\hat{R}$) and effective sample size (ESS).

The BDML yields full posterior distributions over the ATE and ATT, enabling direct probabilistic interpretation of treatment effect uncertainty. Results are reported as posterior means, posterior standard deviations, 95 percent credible intervals, and the posterior probability that the treatment effect exceeds zero. This probability is a model-and-prior-conditional statement, not a claim of logical certainty: a reported value at or near one means that, under the fitted model and the stated weakly informative prior, the posterior allocates negligible mass to zero or negative effects. To confirm that this result is driven by the data rather than the prior, a prior-sensitivity check re-estimates each effect under tighter and more diffuse priors, Normal(0, 1), Normal(0, 2.5), and Normal(0, 5). Posterior means are stable to the third decimal across the three priors, the posterior probability of a positive effect holds at 1.000 throughout, and convergence stays clean ($\hat{R}$ at most 1.0015; effective sample sizes from 4,955 to 5,644). The conclusion is a property of the data, not the prior. Bayesian and frequentist DML estimates are compared as a robustness check on sensitivity to the inferential framework.

5.4 SUTVA Sensitivity Test

The Stable Unit Treatment Value Assumption (SUTVA) requires that an individual household's treatment effect is not influenced by the adoption behaviour of neighbouring households. Given that agricultural extension in Zambia is organised around camp-level service areas, this assumption warrants direct testing (Miguel and Kremer, 2004). The SUTVA sensitivity test augments the PLR covariate matrix with the camp-level leave-one-out mean CSA adoption rate as a spillover proxy and re-estimates the ATE. A materiality threshold of 10 percent shift in the baseline ATE defines the sensitivity boundary: shifts below this threshold are taken as supporting the SUTVA assumption; shifts above it indicate potential interference bias requiring further investigation. The 10 percent threshold is author-specified, chosen ex ante rather than drawn from an established convention; the raw percentage shift is reported alongside it so that readers can apply a stricter standard if preferred.

6. Results

6.1 MESR Estimates: Farm Income

The MESR second-stage outcome equations yield adoption-intensity-specific ATT estimates disaggregated by farm stratum. For farm income the cell-level point estimates are positive across every subsistence regime and across the commercial single- and dual-practice regimes, tracing the gradient

the conceptual framework predicts (Table 1). Among subsistence farmers the first two adoption intensities deliver comparable proportional gains before returns fall at the highest intensity: single-practice adopters (T1) gain ZMW 5,369 (+ 47.1 percent relative to the non-adoption counterfactual, N = 381) and dual-practice adopters (T2) gain ZMW 5,170 (+ 47.2 percent, N = 297), while multi-practice adopters (T3) gain ZMW 1,454 (+ 12.9 percent, N = 233). In absolute terms the subsistence income ATT declines monotonically across intensity ($T1 > T2 > T3$); in proportional terms the single- and dual-practice returns are effectively level at about 47 percent, with the sharp drop reserved for the multi-practice regime. These are the point magnitudes on the Duan-smeared level scale. Their log-scale bootstrap intervals are wide, and none of the six income cell ATTs is individually distinguishable from zero once first-stage selection uncertainty is propagated (Table 1). The precisely estimated income effect, identified on the full sample, is reported by DML and BDML in Section 6.3.

Among commercial farmers, T1 yields ZMW 28,165 (+ 23.1 percent, N = 126), T2 yields ZMW 38,432 (+ 31.4 percent, N = 155), and T3 turns negative at ZMW - 10,422 (-8.5 percent, N = 130). The absolute ATT is larger for commercial farmers in ZMW terms at the lower intensities, consistent with their higher income base. The proportional effect, by contrast, is substantially larger for subsistence farmers at every intensity, consistent with the hypothesis that lower-resource households face steeper marginal returns to initial CSA engagement. As with the subsistence cells, the commercial income ATTs carry wide bootstrap intervals and are not individually significant, so the between-stratum contrast is read as a pattern in the point estimates rather than as a set of separately identified effects.

Across adoption regimes the robust feature in both strata is the fall at multi-practice (T3) intensity. Among subsistence farmers the income ATT declines monotonically in absolute terms and is level across the first two intensities in proportional terms, then drops steeply at T3; among commercial farmers the proportional return rises from T1 to T2 before turning negative at T3. Multi-practice income returns are thus the smallest in the subsistence stratum and negative in the commercial stratum. The pattern points to diminishing marginal returns at the highest intensity of practice stacking within the cross-sectional window captured by the survey. Three explanations are consistent with it: transition costs associated with simultaneous multi-practice adoption, compositional differences among households at each adoption level, and the first one or two practices capturing the bulk of the welfare gain.

Table 1
MESR Average Treatment Effects on the Treated, Farm Income

Stratum	Regime	N	ATT (ZMW)	% Change	Log-scale ATT [95% CI]
Subsistence	T1 vs T0	381	5,369	+ 47.1%	0.51 [-0.49, 1.46]
Subsistence	T2 vs T0	297	5,170	+ 47.2%	0.53 [-0.16, 1.20]
Subsistence	T3 vs T0	233	1,454	+ 12.9%	0.31 [-0.25, 0.94]
Commercial	T1 vs T0	126	28,165	+ 23.1%	0.24 [-1.05, 1.28]
Commercial	T2 vs T0	155	38,432	+ 31.4%	0.27 [-0.87, 1.26]
Commercial	T3 vs T0	130	-10,422	-8.5%	-0.10 [-1.04, 0.74]

Note: Columns report the Duan (1983) smeared level ATT in Kwacha and the percentage change relative to the T0 counterfactual within each stratum, alongside the log-scale switching-regression ATT and its bootstrap 95 percent confidence interval. Inference is on the log-scale ATT, from a full two-stage nonparametric bootstrap (1,000 replicates) that resamples households and re-estimates both the selection and outcome stages on each replicate, so first-stage selection uncertainty enters the standard errors. None of the six income cell intervals excludes zero: the disaggregated income ATTs are positive and follow the expected gradient, but they are imprecisely estimated on the stratum-by-intensity subsamples. The identified income effect for the population and for adopters is reported by DML and BDML (Tables 3 and 5). N refers to households in the relevant adoption regime. Source: Authors' own calculations from WASA Project Survey data.

6.2 MESR Estimates: Crop Yield

The crop yield results reveal a sharp stratum divergence (Table 2). Subsistence yield gains are the one part of the MESR disaggregation that is individually identified: T2 adopters gain ZMW 8,893 per hectare (+ 293.7 percent; log-scale ATT 1.49, 95 percent CI [0.50, 2.29], $p < 0.01$) and T3 adopters gain ZMW 1,558 per hectare (+ 59.5 percent; 0.84, [0.07, 1.67], $p = 0.05$), both with bootstrap intervals that exclude zero, while the large T1 point gain of ZMW 4,488 per hectare (+ 140.1 percent) is positive but imprecisely estimated (1.10, [-0.35, 2.34]). The non-monotonic pattern, with dual-practice (T2) adoption delivering the largest and best-identified subsistence yield gain, suggests that dual-practice adoption may involve specific combinations (for example, conservation agriculture paired with agroforestry) that generate disproportionate yield gains, plausibly through soil-health improvements that lag in their translation to income.

These large percentage gains are read against low non-adoption baselines. Under the Duan-smeared retransformation, the subsistence non-adoption counterfactual yield averages roughly 2,600 to 3,200 ZMW per hectare, specifically about 3,204 for the T1 comparison, 3,028 for T2, and 2,619 for T3, so a large proportional gain reflects a moderate absolute ZMW-per-hectare increase rather than an implausibly large one; the absolute ATTs of ZMW 4,488 and ZMW 8,893 per hectare make this explicit against those bases. The Kwacha-level figures use the Duan (1983) smearing correction rather than naive exponentiation, which removes the low-baseline retransformation bias that would otherwise inflate

the smallest-baseline contrasts; the proportional ranking across intensities, which follows the log-scale coefficients, is unchanged.

Among commercial farmers the yield ATTs are not individually distinguishable from zero at any intensity: the point estimates are small and their bootstrap intervals span zero, and for the single-practice cell the level and log-scale point estimates even differ in sign (ZMW 3,614 per hectare, + 9.7 percent, against a log-scale ATT of -0.11), a reminder that an imprecisely estimated retransformed magnitude should not be read as a directional effect. The negative multi-practice point estimate (ZMW – 4,339 per hectare, -25.3 percent; log-scale ATT – 0.28, 95 percent CI [-1.44, 0.81]) is consistent with a reallocation story, in which commercially oriented multi-practice adopters diversify into higher-value but lower-volume activities and reduce per-hectare staple yield, but the cell is too imprecisely estimated to carry that reading on its own. What the commercial results establish is the absence of a detectable per-hectare yield gain at higher intensity, which alongside the income pattern underscores the importance of evaluating multiple welfare dimensions simultaneously.

Table 2
MESR Average Treatment Effects on the Treated, Crop Yield

Stratum	Regime	N	ATT (ZMW/ha)	% Change	Log-scale ATT [95% CI]
Subsistence	T1 vs T0	381	4,488	+ 140.1%	1.10 [-0.35, 2.34]
Subsistence	T2 vs T0	297	8,893	+ 293.7%	1.49 [0.50, 2.29]
Subsistence	T3 vs T0	233	1,558	+ 59.5%	0.84 [0.07, 1.67]
Commercial	T1 vs T0	126	3,614	+ 9.7%	-0.11 [-1.48, 1.25]
Commercial	T2 vs T0	155	1,335	+ 6.2%	0.19 [-1.27, 1.28]
Commercial	T3 vs T0	130	-4,339	-25.3%	-0.28 [-1.44, 0.81]

Note: Columns are as in Table 1: the Duan (1983) smeared level ATT in Kwacha per hectare and its percentage change, and the log-scale ATT with its bootstrap 95 percent confidence interval. Only two cell intervals exclude zero, both in the subsistence stratum: dual-practice yield (T2, log-scale ATT 1.49, $p < 0.01$) and multi-practice yield (T3, 0.84, $p = 0.05$). The remaining yield ATTs, including every commercial cell, are not individually distinguishable from zero, and for the commercial single-practice cell the level and log-scale point estimates differ in sign, so an imprecisely estimated retransformed magnitude should not be read as a directional effect. The identified population and adopter yield effects are reported by DML and BDML (Tables 3 and 5). N as per Table 1. Source: Authors' own calculations from WASA Project Survey data.

The six strata-by-regime ATTs reported per outcome raise the question of multiple comparisons, but the deeper inferential point is precision. Per-cell inference is placed on the log scale by a full two-stage nonparametric bootstrap that resamples households, refits the selection multinomial logit, rebuilds the Lee (1983) selection terms, and refits the outcome equations on every replicate (1,000 replicates), so first-stage selection uncertainty is carried into the ATT standard errors rather than treated as known.

Under this inference the cell-level ATTs are estimated imprecisely on the small stratum-by-intensity subsamples: only the subsistence yield gains at dual-practice (log-scale ATT 1.49, $p < 0.01$) and multi-practice (0.84, $p = 0.05$) intensity are individually significant, and a Benjamini-Hochberg adjustment within the six-contrast yield family leaves the dual-practice cell significant. The disaggregated estimates are therefore read as a coherent gradient predicted ex ante by the conceptual framework, subsistence gains positive throughout with the largest and best-identified at dual-practice intensity and commercial returns falling toward zero or below at the highest intensity, and not as a set of separately identified coefficients. Identification of the population-average and adopter effects rests on the DML and BDML estimators of Section 6.3, which are estimated on the full sample and are significant at the 1 percent level.

6.3 Double Machine Learning Estimates

The DML PLR model estimates the population ATE of binary CSA adoption on log-transformed outcomes using an analytic sample of $N = 2,016$ households. For farm income, the PLR-ATE is 0.326 (SE = 0.048, $p < 0.001$, 95% CI [0.232, 0.420]), indicating that CSA adoption is associated with approximately a 33 percent increase in farm income after controlling for the full SLF covariate set through machine-learning nuisance estimation (Table 3). For crop yield, the PLR-ATE is 0.468 (SE = 0.067, $p < 0.001$, 95% CI [0.336, 0.600]), corresponding to an approximately 47 percent yield premium across the smallholder population.

The IRM-ATT for farm income is 0.244 (SE = 0.070, $p < 0.001$, 95% CI [0.108, 0.381]), and the IRM-ATT for crop yield is 0.318 (SE = 0.083, $p < 0.001$, 95% CI [0.156, 0.480]). The ATT estimates sit clearly below the ATE estimates in both outcomes, pointing to positive selection into adoption: households that adopted CSA would have achieved somewhat higher outcomes than average even without treatment. Both the PLR and IRM estimates are statistically significant and economically meaningful, and the consistency between the two specifications supports the robustness of the DML identification strategy.

Triangulation with the MESR estimates is directional rather than inferential. The DML IRM-ATT for income (0.244, approximately 24 percent) sits within the span of the MESR income point estimates and matches the pattern of returns falling at the highest adoption intensity, and the DML yield ATE (0.468) and ATT (0.318) align with the positive subsistence yield gains, the best-identified of which, dual-practice, is itself individually significant in the MESR bootstrap. The MESR point estimates therefore corroborate the direction and heterogeneity of the DML effects, while the weight of identification is carried by the agreement between the two full-sample estimators, DML and BDML, whose estimates coincide to within 0.041 and both exclude zero decisively. That convergence of two independently specified full-sample estimators, one frequentist and one Bayesian, provides the robust evidentiary basis for a causal reading of the observed welfare differentials.

Table 3
Double Machine Learning Estimates, ATE and ATT

Outcome	Estimand	Estimate	SE	95% CI	p-value
Farm Income (log)	PLR-ATE	0.326	0.048	[0.232, 0.420]	< 0.001***
Farm Income (log)	IRM-ATT	0.244	0.070	[0.108, 0.381]	< 0.001***
Crop Yield (log)	PLR-ATE	0.468	0.067	[0.336, 0.600]	< 0.001***
Crop Yield (log)	IRM-ATT	0.318	0.083	[0.156, 0.480]	< 0.001***
<i>Note: PLR = Partially Linear Regression (ATE). IRM = Interactive Regression Model (ATT). Nuisance learners: LASSO and Random Forest. 5-fold cross-fitting, set.seed(419). N = 2,016. Standard errors based on the Neyman-orthogonal influence function. Source: Authors' own calculations from WASA Project Survey data.</i>					

6.4 SUTVA Sensitivity Test

The SUTVA sensitivity test augments the PLR covariate matrix with the camp-level leave-one-out mean CSA adoption rate as a proxy for potential spillover interference. For farm income, the augmented ATE is 0.339 (SE = 0.055), a 3.86 percent increase relative to the baseline of 0.326. For crop yield, the augmented ATE is 0.491 (SE = 0.076), a 4.88 percent increase over the baseline of 0.468 (Table 4). Both percentage shifts are computed from the unrounded coefficients and are estimated under the same specification as the main DML results, so the test is internally consistent with Table 3.

Both shifts fall below the pre-specified 10 percent materiality threshold, supporting the SUTVA assumption. The direction of change, upward in both outcomes, indicates that camp-level CSA spillover, if present, would amplify rather than diminish the estimated treatment effect. The baseline estimates are therefore conservative: the household-level ATE under SUTVA understates the true social return to CSA promotion, which would include inter-household spillover benefits transmitted through the camp extension network. The SUTVA assumption is correspondingly satisfied closely enough for valid household-level causal inference.

Table 4
SUTVA Sensitivity Test, Spillover Robustness

Outcome	Baseline ATE	Augmented ATE	% Change	Threshold
Farm Income (log)	0.326	0.339	+ 3.86%	< 10% ✓
Crop Yield (log)	0.468	0.491	+ 4.88%	< 10% ✓
<i>Note: Augmented model adds the camp-level leave-one-out mean adoption rate to the PLR covariate matrix. Materiality threshold: 10 percent change in ATE relative to baseline. Percentage change is computed from unrounded coefficients, so it does not reconcile exactly to the three-decimal ATE columns. Both outcomes within threshold, supporting SUTVA. Source: Authors' own calculations from WASA Project Survey data.</i>				

6.5 Bayesian DML Estimates

Table 5 reports the BDML posterior estimates alongside the frequentist DML point estimates. All parameters converge satisfactorily: $\hat{R}$ values do not exceed 1.001 for any parameter, and effective sample sizes exceed 5,300, indicating adequate posterior sampling throughout.

For farm income, the posterior mean ATE is 0.325 (posterior SD 0.048; 95% CrI [0.231, 0.422]). The income ATT posterior mean is 0.203 (SD 0.032; 95% CrI [0.140, 0.264]). For crop yield, the posterior mean ATE is 0.439 (SD 0.064; 95% CrI [0.312, 0.565]), with the ATT at 0.315 (SD 0.040; 95% CrI [0.239, 0.394]). The posterior probability that the treatment effect exceeds zero is 1.000 for all four estimands, meaning the posterior allocates negligible mass to zero or negative effects under either outcome. This probability is a property of the fitted model and the stated weakly informative prior, and it is reported as corroboration of the frequentist estimates rather than as a statement of certainty; the prior-sensitivity check described in Section 5.3 establishes that the result is data-driven.

Bayesian and frequentist DML estimates agree closely across all four estimands. For the population-average effects they deviate by 0.001 for the income ATE (0.325 versus 0.326) and 0.029 for the yield ATE (0.439 versus 0.468). For the effect on the treated the agreement is tighter still on yield, the posterior mean of 0.315 sitting 0.003 below its frequentist twin of 0.318, and remains close on income, where the posterior mean of 0.203 sits 0.041 below its frequentist twin of 0.244. The largest of the four deviations is 0.041 (income ATT, about 1.3 posterior standard deviations); the other three fall within one posterior standard deviation. Both frameworks keep every estimand positive and far from zero. This cross-framework comparison confirms that the estimates are robust to distributional assumptions.

Figure 3 presents the triangulation visually. For each outcome it places the three estimators on a common log-effect scale: the six MESR regime-by-stratum ATTs of Tables 1 and 2, shown as points after conversion to the log scale through $\ln(1 + \text{ATT percent} / 100)$; the two DML estimates of Table 3 (PLR-ATE and IRM-ATT) with their 95 percent confidence intervals; and the two Bayesian DML posterior means of Table 5 (ATE and ATT) with their 95 percent credible intervals. The DML and BDML ATE estimates cluster tightly for each outcome, while the MESR regime ATTs scatter around them, the subsistence-yield ATTs sitting furthest to the right because they are computed against the very low non-adoption baselines discussed in Section 6.2. That clustering of the two population-level estimators, with the MESR regime effects distributed around them, is the visual statement of the triangulation. Every value is taken directly from Tables 1 to 5; none is newly generated. The MESR points are shown without individual error bars to keep the figure legible; their per-cell bootstrap 95 percent confidence intervals are reported in Tables 1 and 2 and, on the small stratum-by-intensity subsamples, are wide, so only the subsistence yield dual- and multi-practice intervals exclude zero.

Table 5
Bayesian DML Estimates, Farm Income and Crop Yield (N = 2,016)

Outcome	Estimand	Post. mean	Post. SD	95% CrI	P(> 0)	R	ESS	Freq.
Income (log)	ATE	0.325	0.048	[0.231, 0.422]	1.000	1.001	5,656	0.326
Income (log)	ATT	0.203	0.032	[0.140, 0.264]	1.000	1.000	5,346	0.244
Yield (log)	ATE	0.439	0.064	[0.312, 0.565]	1.000	1.001	5,714	0.468
Yield (log)	ATT	0.315	0.040	[0.239, 0.394]	1.000	1.001	5,634	0.318

Note: Post. mean = posterior mean; Post. SD = posterior standard deviation; 95% CrI = 95 percent credible interval; P(> 0) = posterior probability that the effect exceeds zero, returned as 1.000 by the sampler; R = Gelman-Rubin convergence diagnostic; ESS = effective sample size; Freq. = frequentist DML estimate for comparison (from Table 3). Source: Authors' own calculations from WASA Project Survey data.

7. Discussion

The results carry four principal findings of theoretical and policy relevance. CSA adoption raises smallholder welfare: the population-average and adopter effects on both income and yield are positive and significant at the 1 percent level under DML and are corroborated by the Bayesian posterior, which places the probability of a positive effect at 1.000. The MESR disaggregation traces the same effect across farm strata and adoption intensities as a consistent positive gradient, individually significant for the subsistence yield gains and directionally consistent elsewhere, confirming the conclusions of prior switching regression studies (Teklewold et al., 2013) in a new country and methodological context. The convergence of two independently specified full-sample estimators, frequentist DML and Bayesian DML, with the direction and heterogeneity of the MESR point estimates, provides stronger evidentiary grounds for a causal interpretation than any single estimator could support.

These income and yield gains carry an indirect but concrete food-security reading. For a smallholder household that both produces and purchases food, the approximately 33 percent income premium widens the budget available to acquire a more diverse and more stable diet, addressing the access dimension of food security, while the yield gains, concentrated among subsistence farmers, raise the volume of own-produced staples available for home consumption, addressing the availability dimension (Kassie et al., 2015; Kitsuki and Sakurai, 2023). The food-consumption module captured in the WASA instrument, meals-per-day and food-frequency by group, would allow this pathway to be tested directly, and that companion analysis is the natural next step; the present estimates establish the upstream causal links on which any such food-security effect would depend.

Among subsistence farmers the income return is level across the first two adoption intensities, about 47 percent at both single- and dual-practice adoption, and falls sharply at multi-practice (T3) intensity; among commercial farmers it rises to T2 before turning negative at T3. The consistent feature across both strata, visible in the point estimates though not sharply resolved cell by cell, is the fall at the highest intensity: multi-practice income returns are the smallest of the three regimes for subsistence farmers and turn negative for commercial farmers. Read together with the DML finding that the effect on the treated sits below the population-average effect, a selection signature pointing the same way, this pattern challenges programme designs that prioritise the broadest multi-practice adoption as the preferred end-state. The data indicate that welfare returns concentrate in the first one or two practices adopted and thin out as the portfolio widens to three or more, implying that dissemination of one or two well-chosen practices may generate greater aggregate welfare, and greater aggregate food-access gains, than concentrated promotion of the widest multi-practice portfolios.

The stratum divergence in yield effects, subsistence farmers realising individually identified gains while commercial farmers show no detectable per-hectare gain and a negative but imprecisely estimated point estimate at multi-practice intensity, is consistent with a reallocation mechanism specific to commercially oriented producers at the highest intensity, though the commercial cells are too imprecise to confirm it. This pattern has direct implications for targeting: CSA programmes designed around yield maximisation may be suboptimal for commercial smallholders who are already diversifying into higher-value activities. Programme design should accommodate heterogeneous objectives and production systems rather than applying uniform practice mandates. From a food-security standpoint, the subsistence concentration of the yield gains is favourable, since it is the subsistence households whose own-production most directly governs their food availability.

The SUTVA sensitivity test reveals that camp-level spillover effects are present but modest, falling below the materiality threshold in both outcomes. The conservative direction of the bias, spillover would amplify the true social return, strengthens the case for community-level or camp-level programme evaluation as a complement to household-level impact assessment. This result is consistent with social learning and information diffusion mechanisms identified by Conley and Udry (2010) and indicates that future studies should model spillover explicitly rather than assuming it away.

A further implication concerns the timing of vulnerability along the adoption pathway. The concentration of income returns in the first two practices, and their fall at the third, identifies the early adoption transitions as the welfare-critical moments in the CSA diffusion pathway, and also the moments of maximum financial vulnerability for the smallholder. A single adverse shock in the season immediately following initial adoption, a rainfall deficit, an input price spike, or a pest infestation, can eliminate the income gain and extinguish the farmer's willingness to re-adopt. The empirical reversion risk is documented in the adoption literature (Feder et al., 1985; Teklewold et al., 2013), and the present findings locate the first adoption transitions as where reversion is most likely and most consequential, since it can permanently remove the farmer from the adopting population, and with that loss removes the food-security gains the transition would have delivered. Protecting the early adopter against

catastrophic loss, through water security, soil resilience, and financial protection, converts the welfare gain from a fragile one-season event into a repeating livelihood trajectory; this is a productive avenue for public-private partnership design and future research.

The present analysis operationalises CSA adoption as an ordered intensity count, the number of distinct bundle categories adopted, rather than as a vector of specific bundle combinations. This design supports the primary research questions but does not directly identify which specific combination of categories produces the greatest welfare gain for a first-time adopter. Addressing this rigorously would require bundle-specific switching regression models estimated separately for each of the six Amadu et al. (2020) categories, a productive avenue for future research. The available evidence nonetheless supports a preliminary combination logic: non-woody cultivation practices, intercropping, cover cropping, and legume integration, rank among the most widely adopted CSA practices in the Zambian smallholder context (Arslan et al., 2014; Ng'ombe et al., 2017) and require minimal capital, making them an accessible entry point for a T0 farmer; complementing these with residue addition builds a soil-improving bundle addressing productivity and resilience together. This combination recommendation extends beyond the formal causal estimates reported here, which establish intensity effects rather than combination-specific effects, and the warrant for it is reserved for the bundle-specific research identified above.

8. Conclusion and Policy Implications

This study provides causal evidence on the welfare effects of CSA adoption among smallholder farming households in Zambia, employing a methodological triangulation of Multinomial Endogenous Switching Regression, Double Machine Learning, and Bayesian Double Machine Learning. The SUTVA sensitivity test confirms the validity of household-level treatment effect estimates. Three policy implications follow, each with a food-security dimension given that income and own-production are the upstream determinants of household food access and availability.

CSA promotion is welfare-improving for Zambian smallholders: the population-average and adopter effects on income and yield are positive and significant across the DML and BDML frameworks, and the MESR point estimates trace the same positive gradient across farm types and adoption regimes, so programme investment in CSA dissemination is justified on causal grounds. The strong proportional returns to early adoption among subsistence farmers indicate that first-mover dissemination, getting non-adopters to adopt one or two practices, should be the priority for welfare and food-access targeting.

The fall in returns at the highest adoption intensity cautions against multi-practice mandates as a universal programme objective. Both strata show their smallest income point returns at the multi-practice (T3) regime, and the commercial multi-practice point estimates turn negative in both income and yield, though these cell-level differences are not individually significant, so extension resources may be more effectively deployed in expanding the breadth of single- and dual-practice adoption across non-adopting households than in deepening multi-practice portfolios among already-adopting households.

This recommendation is conditional on the cross-sectional data window; longitudinal evidence may reveal stronger long-run returns to practice stacking that a single survey round cannot show.

The camp-level spillover evidence, although below the materiality threshold, supports investment in group-based and camp-level CSA learning platforms that leverage peer-to-peer information diffusion. The positive direction of the spillover bias implies that the social return to CSA promotion exceeds the private return measured at the household level, strengthening the case for public subsidy of extension and demonstration programmes.

This study has limitations that bound the scope of inference. The data are cross-sectional, preventing panel difference-in-differences strategies that would further reduce omitted variable concerns. The conditional independence assumption underlying the DML estimator, though supported by an extensive covariate matrix spanning the five capital dimensions, cannot be formally tested; the unconfoundedness sensitivity analysis in Section 5.2 bounds, rather than eliminates, this exposure. The welfare outcomes analysed are income and yield, which proxy the access and availability dimensions of food security but do not measure dietary diversity or consumption directly; the WASA food-consumption module supports a companion analysis that would close this gap. The Kwacha-level MESR effects are recovered using the Duan (1983) smearing retransformation rather than naive exponentiation, which removes the low-baseline retransformation bias; the log-scale coefficients, and thus the qualitative findings, do not depend on this choice. The MESR disaggregation is itself limited in precision: on the small stratum-by-intensity subsamples, and once first-stage selection uncertainty is propagated through the two-stage bootstrap, only the subsistence yield gains at dual- and multi-practice intensity are individually significant, so the disaggregated ATTs are interpreted as a heterogeneity pattern while identification of the welfare effects rests on the full-sample DML and BDML estimators. Future research should pursue longitudinal follow-up of the WASA Project cohort to estimate dynamic adoption effects and long-run welfare and food-security trajectories.

Declarations

Funding. Data collection under the Water and Soil Accelerator (WASA) Project was funded by the United States Agency for International Development (USAID). No separate funding was received for the analysis reported in this article.

Competing interests. The authors declare no competing interests.

Ethics approval. This study analyses secondary, de-identified household-survey data collected under the WASA Project. Ethics clearance was granted by the Faculty of Natural and Agricultural Sciences Ethics Committee (FNAS-REC) of North-West University (reference NWU-00551-25-A9; approved 21 November 2025), under the authority of the North-West University Senate Committee for Research Ethics.

Consent to participate. Informed consent was obtained from all households included in the WASA Project survey at the point of data collection.

Consent for publication. Not applicable. No individual person's identifying data are reported; all outputs are aggregated.

Data availability. The household-level survey data are confidential and not publicly available, owing to the sensitivity of the information and the ethics conditions under which they were collected. De-identified data may be available from the corresponding author on reasonable request, subject to North-West University ethics approval and WASA Project permission.

Code availability. The analysis code (R; DoubleML and brms; set.seed(419) for reproducibility) is available from the corresponding author on reasonable request.

Authors' contributions. E.K. Chinembiri (lead author): conceptualisation, writing of the original draft, and project administration. M. Mutenje: conceptualisation, methodology, formal analysis, investigation (data collection), and manuscript review and editing. N.D. Malahlela: supervision.

ORCID.

Evans K. Chinembiri: <https://orcid.org/0009-0003-7304-0214>

Munyaradzi Mutenje: <https://orcid.org/0000-0002-7829-9300>

Nkele D. Malahlela: <https://orcid.org/0009-0001-1653-7553>

Acknowledgements. The authors thank the Water and Soil Accelerator (WASA) Project, the 27 enumerators who conducted the fieldwork, and the participating smallholder households.

Notes

¹ A penalised (ridge-regularised) multinomial refit under set.seed(419), recorded in the post-estimation pipeline, confirms that the crop adoption rate's association with adoption intensity is large, positive, and monotone, with first-stage coefficients of approximately 2.0, 3.0, and 3.6 in the T1, T2, and T3 equations and point estimates stable across a range of penalty strengths. Even under penalisation, however, the Hessian-based per-regime standard errors remain numerically degenerate, so they are not reported.

References

1. Amadu FO, McNamara PE, Miller DC (2020) Understanding the adoption of climate-smart agriculture: a farm-level typology with empirical evidence from southern Malawi. *World Dev* 126:104692. <https://doi.org/10.1016/j.worlddev.2019.104692>
2. Arslan A, McCarthy N, Lipper L, Asfaw S, Cattaneo A (2014) Adoption and intensity of adoption of conservation farming practices in Zambia. *Agric Ecosyst Environ* 187:72–86. <https://doi.org/10.1016/j.agee.2013.08.017>

3. Asfaw S, Shiferaw B, Simtowe F, Lipper L (2012) Impact of modern agricultural technologies on smallholder welfare: evidence from Tanzania and Ethiopia. *Food Policy* 37(3):283–295. <https://doi.org/10.1016/j.foodpol.2012.02.013>
4. Bach P, Chernozhukov V, Kurz MS, Spindler M, Klaassen S (2024) DoubleML: an object-oriented implementation of double machine learning in R. *J Stat Softw* 108(3):1–56. <https://doi.org/10.18637/jss.v108.i03>
5. Bourguignon F, Fournier M, Gurgand M (2007) Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons. *J Economic Surveys* 21(1):174–205. <https://doi.org/10.1111/j.1467-6419.2007.00503.x>
6. Carletto C, Zezza A, Banerjee R (2013) Towards better measurement of household food security: harmonizing indicators and the role of household surveys. *Global Food Secur* 2(1):30–40. <https://doi.org/10.1016/j.gfs.2012.11.006>
7. Chambers R, Conway G (1992) Sustainable rural livelihoods: practical concepts for the 21st century. IDS Discussion Paper 296. Institute of Development Studies, Brighton
8. Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J (2018) Double/debiased machine learning for treatment and structural parameters. *Economet J* 21(1):C1–C68. <https://doi.org/10.1111/ectj.12097>
9. Cinelli C, Hazlett C (2020) Making sense of sensitivity: extending omitted variable bias. *J Royal Stat Society: Ser B (Statistical Methodology)* 82(1):39–67. <https://doi.org/10.1111/rssb.12348>
10. Conley TG, Udry CR (2010) Learning about a new technology: pineapple in Ghana. *Am Econ Rev* 100(1):35–69. <https://doi.org/10.1257/aer.100.1.35>
11. DiTraglia FJ, Liu L (2025) Bayesian double machine learning for causal inference. *arXiv:2508.12688*. <https://arxiv.org/abs/2508.12688>
12. Duan N (1983) Smearing estimate: a nonparametric retransformation method. *J Am Stat Assoc* 78(383):605–610. <https://doi.org/10.1080/01621459.1983.10478017>
13. FAO (2009) Declaration of the World Summit on Food Security. World Summit on Food Security, Rome, 16–18 November 2009, WSFS 2009/2. Food and Agriculture Organization of the United Nations, Rome
14. Feder G, Just RE, Zilberman D (1985) Adoption of agricultural innovations in developing countries: a survey. *Econ Dev Cult Change* 33(2):255–298. <https://doi.org/10.1086/451461>
15. Kassie M, Teklewold H, Marenja P, Jaleta M, Erenstein O (2015) Production risks and food security under alternative technology choices in Malawi: application of a multinomial endogenous switching regression. *J Agric Econ* 66(3):640–659. <https://doi.org/10.1111/1477-9552.12099>
16. Kitsuki A, Sakurai T (2023) Vulnerability to seasonal food insecurity as an exposure to risk: the case of the Southern Province of Zambia. *Agric Food Secur* 12(1). <https://doi.org/10.1186/s40066-023-00442-4>
17. Lee LF (1983) Generalized econometric models with selectivity. *Econometrica* 51(2):507–512. <https://doi.org/10.2307/1912003>

18. Lipper L, Thornton P, Campbell BM, Baedeker T, Braimoh A, Bwalya M, Caron P, Cattaneo A, Garrity D, Henry K, Hottle R, Jackson L, Jarvis A, Kossam F, Mann W, McCarthy N, Meybeck A, Neufeldt H, Remington T, Sen PT, Sessa R, Shula R, Tibu A, Torquebiau EF (2014) Climate-smart agriculture for food security. *Nat Clim Change* 4(12):1068–1072. <https://doi.org/10.1038/nclimate2437>
19. Manda J, Alene AD, Mukuma C, Chikoye D (2017) Ex-ante welfare impacts of adopting maize-soybean rotation in eastern Zambia. *Agric Ecosyst Environ* 249:22–30. <https://doi.org/10.1016/j.agee.2017.07.030>
20. Miguel E, Kremer M (2004) Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72(1):159–217. <https://doi.org/10.1111/j.1468-0262.2004.00481.x>
21. Ng'ombe JN, Kalinda TH, Tembo G (2017) Does adoption of conservation farming practices result in increased crop revenue? Evidence from Zambia. *Agrekon* 56(2):205–221. <https://doi.org/10.1080/03031853.2017.1312467>
22. Nhemachena C, Hassan R (2007) Micro-level analysis of farmers' adaptation to climate change in Southern Africa. IFPRI Discussion Paper No. 00714. International Food Policy Research Institute, Washington, DC
23. Ogisi OD, Begho T (2023) Adoption of climate-smart agricultural practices in sub-Saharan Africa: a review of the progress, barriers, gender differences and recommendations. *Farming Syst* 1(2):100019. <https://doi.org/10.1016/j.farsys.2023.100019>
24. Oster E (2019) Unobservable selection and coefficient stability: theory and evidence. *J Bus Economic Stat* 37(2):187–204. <https://doi.org/10.1080/07350015.2016.1227711>
25. Robinson PM (1988) Root-N-consistent semiparametric regression. *Econometrica* 56(4):931–954. <https://doi.org/10.2307/1912705>
26. Rogers EM (2003) *Diffusion of innovations*, 5th edn. Free, New York
27. Sitko N (2007) Maize, food insecurity, and the field of performance in southern Zambia. *Agric Hum Values* 25(1):3–11. <https://doi.org/10.1007/s10460-007-9075-z>
28. Teklewold H, Kassie M, Shiferaw B (2013) Adoption of multiple sustainable agricultural practices in rural Ethiopia. *J Agric Econ* 64(3):597–623. <https://doi.org/10.1111/1477-9552.12011>

Figures

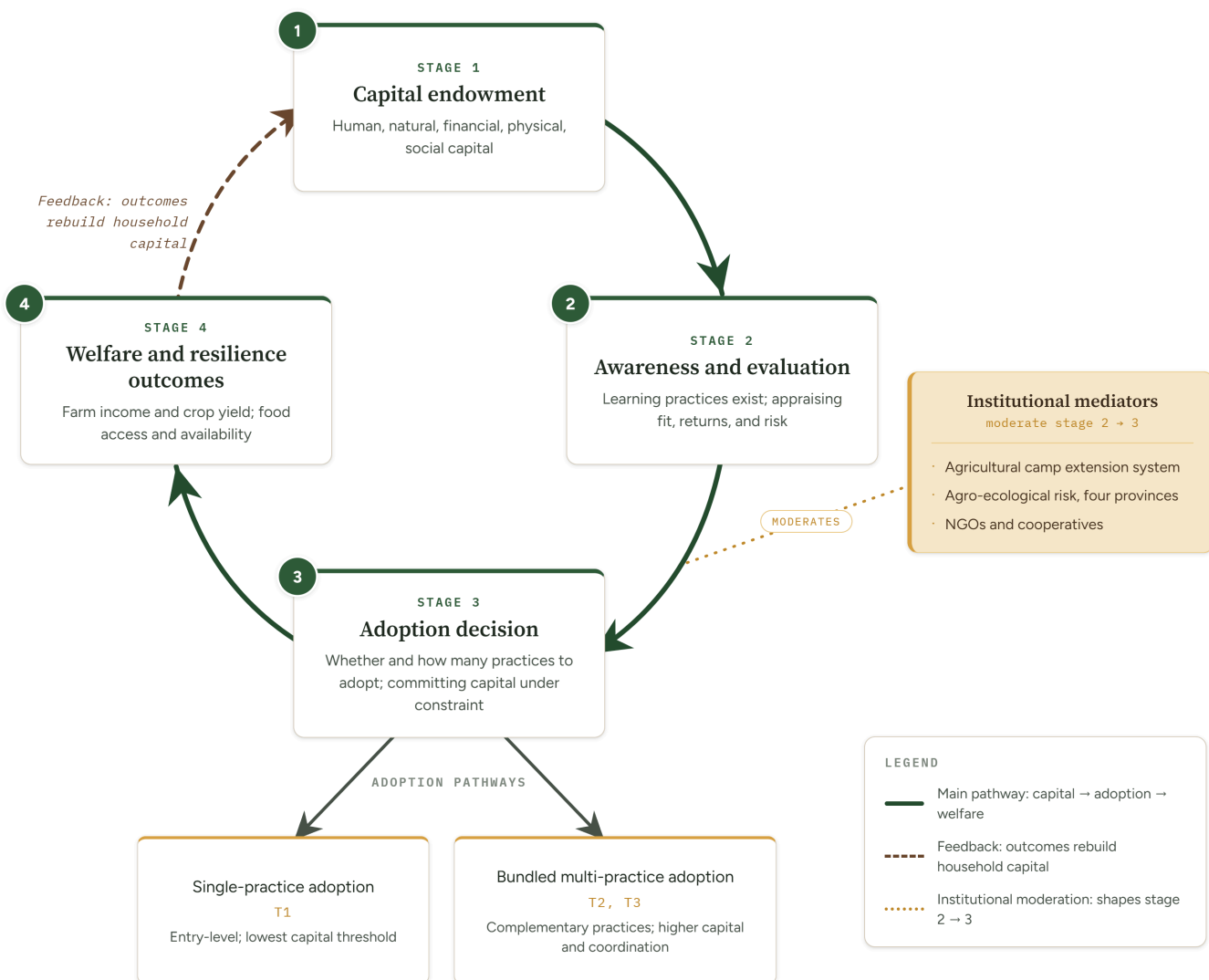


Figure 1

Conceptual framework: household capital to welfare through CSA adoption. The four stages form a circular process with a capital-rebuilding feedback loop from stage four to stage one; Zambia-specific institutional mediators moderate the stage-two-to-stage-three pathway. The framework adapts the Sustainable Livelihoods Framework (Chambers and Conway, 1992), the diffusion-of-innovations tradition (Rogers, 2003), and the agricultural technology adoption literature (Feder et al., 1985).

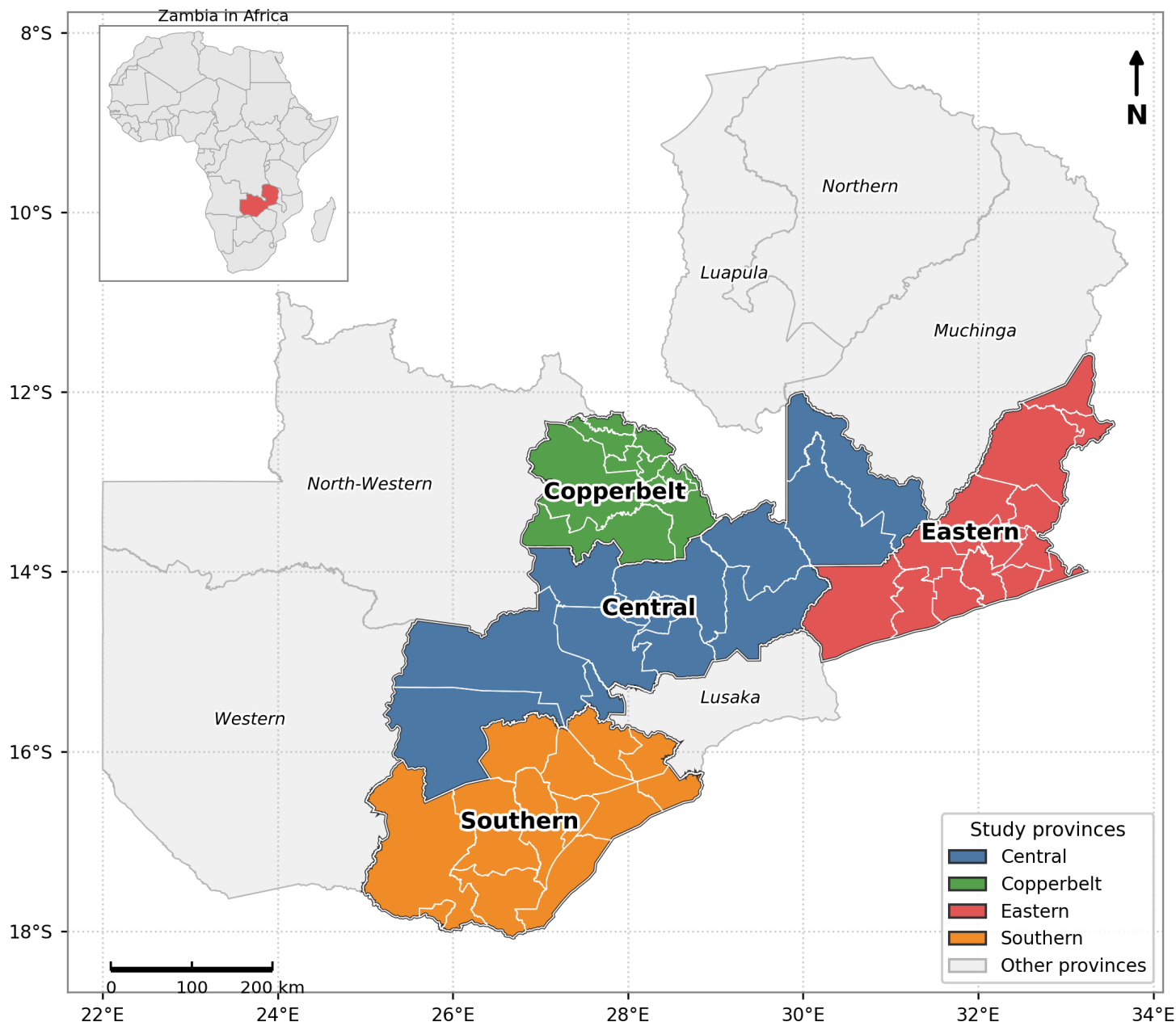


Figure 2

Study area. The four provinces covered by the WASA Project household survey, Central, Copperbelt, Eastern, and Southern, are shown in distinct colours with their internal district boundaries; the remaining six provinces are muted. The inset locates Zambia within Africa. Province and district boundaries are drawn from the geoBoundaries open administrative dataset (admin-1 and admin-2). Source: authors' construction from WASA Project Survey sampling records.

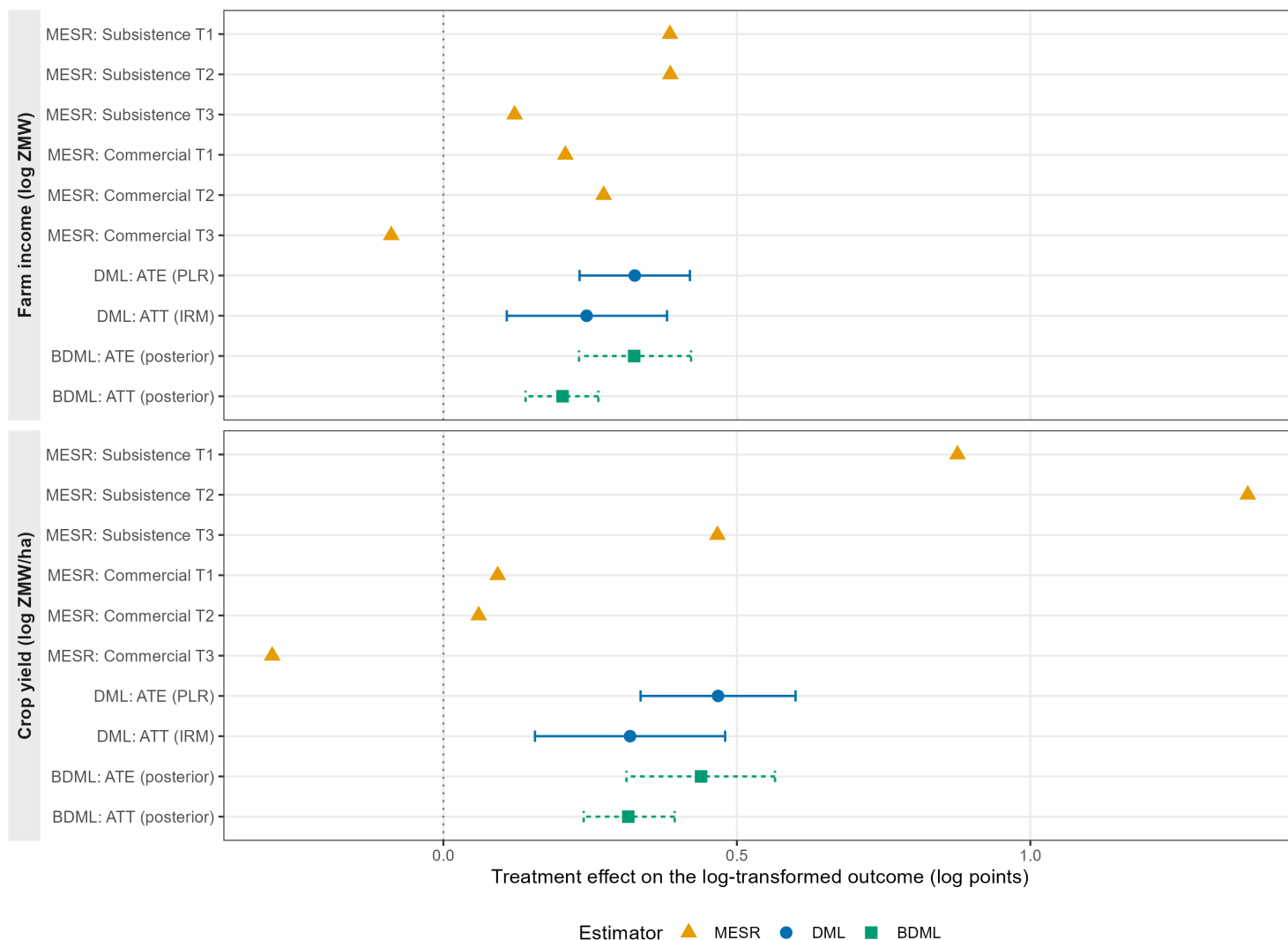


Figure 3

Triangulation of CSA welfare effects across three estimators. Points are the MESR regime-by-stratum ATTs (Tables 1 and 2) on the log scale; horizontal bars are the DML 95 percent confidence intervals (Table 3, solid) and the Bayesian DML 95 percent credible intervals (Table 5, dashed). The dotted vertical line marks the null. Estimators are distinguished by marker shape so the figure reads in greyscale. Source: authors' calculations from WASA Project Survey data.