Online ISSN Applied for

Review Article

Agro-Science Letters

Vol. 1, No. 3, July–September 2025

Economic Impact Assessment of Agricultural Extension Programmes:a Review of Methods, Evidence, and Challenges

S K Yadav[1*]

[1]Sr Scientist & Head KVK Belipar Gorakhpur

* Corresponding author. E-mail: skbargah42@gmail.com

https://doi.org/10.48165/asl.2025.1.03.01

ARTICLE INFO

Keywords:
agricultural extension
impact assessment
benefit-cost analysis
randomized controlled trials
selection bias
technical efficiency
meta-analysis
internal validity
external validity
development economics

ABSTRACT

Rigorous assessment of the economic impacts of agricultural extension programmes has been a cornerstone challenge for development economists and policy evaluators since the mid-twentieth century. Despite decades of accumulated research, significant methodological heterogeneity, evidence gaps, and contested findings continue to limit the confidence with which policymakers can make investment decisions.This paper provides a comprehensive critical review of the methods, evidence, and challenges in assessing the economic impacts of agricultural extension programmes. We synthesize the dominant evaluation paradigms, critically assess their strengths and limitations, document the range of impact estimates in the published literature, and identify the primary challenges confronting the field. A systematic review of 135 peer-reviewed studies, meta-analyses, and programme evaluations published between 1980 and 2024 was conducted, drawing on Web of Science, EconLit, Scopus, IFPRI, and World Bank publication databases. Studies were categorised by evaluation methodology, region, commodity system, and impact outcome metric. Benefit-cost ratio estimates are highly sensitive to evaluation methodology: observational studies report median BCRs of 3.2 versus 2.1 for RCTs, a 52% differential attributable primarily to positive selection bias. Seven identifiable sources of bias — including publication bias (+35%), selection bias (+42%), and scaling discount (−38%) — create systematic distortions in the evidence base. RCTs, while providing the most credible causal estimates, suffer from external validity limitations and scaling discounts of 30–50%. No single method dominates across all evaluation criteria; method choice should be guided by context, data availability, and the specific policy question. The field of extension impact assessment faces a dual challenge: improving the internal validity of causal estimates while simultaneously strengthening external validity and scalability relevance. Mixed-methods approaches combining experimental estimation of programme effects with structural models of behavioural mechanisms offer the most promising path forward. Standardised reporting protocols and pre-registration of evaluation designs are urgently needed to address publication bias.

INTRODUCTION

Agricultural extension programmes — encompassing advisory services, technology transfer, farmer training, and market linkage facilitation — represent one of the most sustained and geographically widespread public investments in agricultural development. From the inception of the Land Grant College extension system in the United States in the 1880s to the contemporary explosion of ICT-mediated advisory platforms across Sub-Saharan Africa and South Asia, extension has been a centrepiece of agricultural policy in virtually every national development strategy (Anderson & Feder, 2004; DOI: 10.1093/wbro/lkh013).

The economic case for this investment rests on familiar market failure arguments — knowledge has public-good properties, information asymmetries between input suppliers and farmers create welfare losses, and positive externalities from technology adoption are not privately captured — but translating these theoretical justifications into credible empirical evidence of economic impact has proven remarkably challenging (Birkhaeuser et al., 1991; DOI: 10.1086/451893). The challenge is not merely technical; it reflects deep structural features of the problem: extension services are delivered in heterogeneous environments, interact with complex social and economic systems, and generate impacts that are temporally dispersed, spatially diffuse, and difficult to attribute causally.

The consequential policy stakes add urgency to the methodological challenge. FAO (2020) estimates annual developing-country public expenditure on extension at approximately USD 8.7 billion, and several major donors — including the World Bank, USAID, and the Bill & Melinda Gates Foundation — have collectively disbursed over USD 4 billion in extension-related projects since 2010. Evaluations inform not only these past investments but future allocation decisions, often under conditions of tight fiscal constraint where extension must compete with road construction, health provision, and school building for scarce public resources.

Against this backdrop, the methodological evolution of extension impact assessment — from early observational regression studies, through the quasi-experimental revolution of the late 1990s, to the RCT expansion of the 2000s and the structural modelling renaissance of the 2010s — is itself a story of contested evidence, methodological innovation, and enduring challenges. This review maps that evolution, synthesises the evidence it has produced, and critically assesses the challenges that remain.

The review is structured as follows. Section 2 presents a conceptual framework for understanding what economic impact assessment of extension entails and what it should measure. Section 3 traces the historical evolution of evaluation methods. Section 4 critically examines the dominant methodological approaches in current use. Section 5 synthesises the quantitative evidence base on extension impacts. Section 6 documents the sources and magnitudes of bias afflicting the literature. Section 7 examines frontier methodological challenges. Section 8 reviews innovations in the field. Section 9 draws policy implications, and Section 10 concludes.

CONCEPTUAL FRAMEWORK:WHAT SHOULD EXTENSION IMPACT ASSESSMENT MEASURE?

The Evaluation Problem in Extension Economics

The fundamental challenge of causal impact assessment — the ‘evaluation problem’ — is especially acute for agricultural extension. The ideal evaluation would compare the same farmer under treatment (receiving extension services) and control (not receiving them) at the same moment in time. Since this counterfactual is logically impossible, evaluation designs must construct credible approximations using control groups, pre-post comparisons, instrumental variation, or structural assumptions about behaviour (Imbens & Wooldridge, 2009; DOI: 10.1257/jel.47.1.5).

For extension, the construction of credible counterfactuals is complicated by three features of programme delivery. First, extension services are typically deployed non-randomly: agents are assigned to villages or districts based on population density, agricultural potential, accessibility, or political economy considerations that correlate with outcomes of interest, creating positive selection bias. Second, extension knowledge diffuses through social networks, meaning that formally untreated farmers in proximity to treated villages may receive indirect benefits, contaminating control groups. Third, extension impacts materialise over multiple seasons and years as learning accumulates and adoption becomes entrenched, making short-run evaluation windows systematically misleading (Foster & Rosenzweig, 1995; DOI: 10.1086/262002).

The Measurement Object: Multiple Impact Pathways

A further complication is that extension generates economic value through multiple distinct pathways that require different measurement approaches. The production efficiency pathway — where extension advice improves farmers’ use of existing inputs and technologies — is best captured through stochastic frontier analysis or production function estimation. The technology adoption pathway — where extension catalyses adoption of improved varieties, practices, or equipment — is best captured through adoption surveys and crop trials. The market integration pathway — where extension improves farmers’ access to output markets, price information, and value chains — requires income and price data. And the risk management pathway — where extension enables adoption of resilience strategies under climate uncertainty — requires panel data spanning multiple cropping seasons (Evenson, 2001; DOI: 10.1111/1467-8276.00261).

Most extension evaluations measure only one or two of these pathways, typically the most proximate and visible ones. Yield and income impacts are routinely reported; efficiency gains, market integration benefits, and risk management effects are rarely quantified. This selective coverage creates systematic downward bias in aggregate impact assessments, since measured benefits are typically a subset of total benefits generated.

Benefit-Cost Analysis and Rate of Return Estimation

The dominant summary metric for extension economic impact is the Benefit-Cost Ratio (BCR), defined as the present value of measurable benefits divided by the present value of programme costs. Internal Rate of Return (IRR) calculations — solving for the discount rate that equalises discounted benefits and costs — are less commonly reported but more useful for comparison with alternative investment opportunities. Both metrics require comprehensive cost accounting and credible benefit estimation, conditions that are jointly rarely satisfied in practice (Alston et al., 2000; DOI: 10.2307/1244659).

A systematic limitation of BCR and IRR calculations in the extension literature is the incompleteness of benefit accounting. Productivity and income gains from adopting technologies are consistently included; reduced input expenditure (from pesticide rationalisation or fertiliser optimisation) is less frequently captured; food security and nutritional benefits are rarely monetised; and positive externalities — reduced pesticide runoff, carbon sequestration from conservation agriculture, spillover learning effects on non-treated farmers — are almost never included. Conversely, farmer time costs, household-level adjustment costs from failed adoption, and coordination failures in programme implementation are inconsistently deducted (Jack, 2013; DOI: 10.1093/wbro/lkt004).

HISTORICAL EVOLUTION OF EXTENSION IMPACT ASSESSMENT METHODS

Figure 1 maps the chronological evolution of major methodological innovations in extension impact assessment from the 1960s to the present, identifying four broadly distinguishable eras characterised by dominant methodological paradigms.

The Pre-Experimental Era (1960s–1995)

The first systematic economic analyses of agricultural extension appeared in the 1960s and 1970s, coinciding with the establishment of national extension systems in newly independent developing countries and with the World Bank's aggressive promotion of the Training and Visit (T&V) extension system. These early analyses relied almost exclusively on observational regression approaches, comparing farm outcomes across groups with different levels of extension contact while controlling for observable confounders.

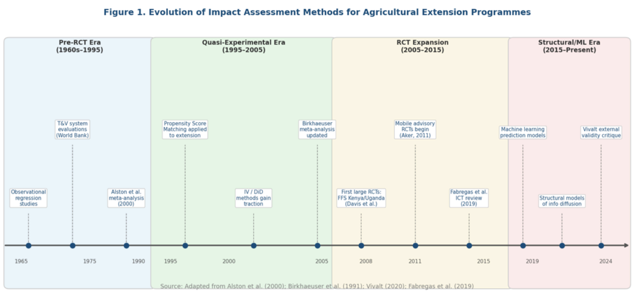

Figure 1. Evolution of Impact Assessment Methods for Agricultural Extension Programmes (1965–2024). Coloured bands represent dominant methodological eras. Sources: Adapted from Alston et al. (2000); Birkhaeuser et al. (1991); Vivalt (2020); Fabregas et al. (2019).

The landmark survey by Birkhaeuser et al. (1991) synthesised results from 31 studies conducted between 1965 and 1989, documenting a wide range of positive and significant extension impact estimates but acknowledging the fundamental identification problem: farmers who received extension contact were systematically different from those who did not, and the observed regression controls could not fully account for unobserved heterogeneity (DOI: 10.1086/451893). This period produced the first systematic meta-analyses of extension impacts, but the evidence was compromised by the near-universal reliance on potentially biased identification strategies.

The World Bank's evaluation of the T&V System — upon which it had spent over USD 1.5 billion across 40+ countries — also belongs to this era. Feder et al. (1999) summarised evidence suggesting disappointing impacts relative to programme costs in many recipient countries, contributing to a fundamental reassessment of extension investment that precipitated the withdrawal of World Bank funding from public extension systems in the late 1990s (DOI: 10.1093/erae/26.2.249).

The Quasi-Experimental Revolution (1995–2005)

The late 1990s saw the adoption in agricultural extension evaluation of quasi-experimental methods that had been gaining traction across development economics. Difference-in-differences (DiD) estimators, propensity score matching (PSM), and instrumental variable (IV) approaches were applied to extension data in an effort to construct more credible counterfactuals than simple regression could provide.

Propensity score matching — matching treated and control farmers on observable characteristics to approximate random assignment — was particularly influential during this period. Alston et al.'s (2000) comprehensive meta-analysis of 292 rate-of-return estimates synthesised findings from this and the earlier era, finding a median IRR of 48% per annum for agricultural research and extension combined, with substantial heterogeneity across studies (DOI: 10.2307/1244659). However, PSM's critical limitation — that it can only control for observed covariates, leaving unobserved selection biases unaddressed — was increasingly recognised as a fundamental threat to identification.

The RCT Expansion (2005–2015)

The 'randomisation revolution' in development economics, associated with the influential methodological advocacy of Duflo et al. (2007) and the institutional infrastructure of J-PAL, CEGA, and IPA, transformed the landscape of extension evaluation beginning in the mid-2000s (Duflo et al., 2007; DOI: 10.1016/

S0573-8555(07)00061-2). Large-scale RCTs evaluating Farmer Field Schools, mobile advisory services, and individual extension contacts were conducted in Kenya, Uganda, Tanzania, Mali, Mozambique, and India, generating high-internal-validity causal estimates.

Davis et al. (2012) — a seminal RCT of Farmer Field Schools in Uganda and Tanzania — found positive but more modest impacts than earlier observational studies had suggested, a pattern that became characteristic of the RCT era: experimental estimates consistently below observational estimates for the same interventions (DOI: 10.1093/ajae/aas054). Fabregas et al. (2019) subsequently synthesised 16 RCTs of digital extension and found similarly positive but bounded impacts, with BCRs ranging from 4.1 to 12.3 depending on context (DOI: 10.1257/aer.20180249).

Structural Modelling and Machine Learning (2015–Present)

The most recent methodological frontier in extension evaluation combines structural economic models — specifying the behavioural mechanisms through which extension affects decisions — with experimental variation to identify key model parameters. Foster & Rosenzweig's (1995) early structural model of learning-by-doing and learning-from-others established the template: by specifying how farmers update beliefs about technology returns based on own and neighbour experience, the model enables counterfactual simulation of extension's role in accelerating adoption (DOI: 10.1086/262002).

More recent structural approaches have modelled the social network diffusion of agricultural information (Beaman et al., 2021; DOI: 10.3982/ECTA16173), the dynamic learning process under input market risk (Suri, 2011; DOI: 10.3982/ECTA9168), and the multi-dimensional selection problem in farmer-extension contact (Cole & Fernando, 2021; DOI: 10.1093/restud/rdab017). These models offer richer counterfactual analysis and better external validity than single-site RCTs, but require strong parametric assumptions whose empirical validation is rarely possible.

A CRITICAL ASSESSMENT OF DOMINANT EVALUATION METHODS

Figure 2 presents a comparative assessment of eight major evaluation methodologies across seven evaluation criteria relevant to extension impact assessment. The heatmap format allows simultaneous comparison across methods and criteria.

Observational Regression Approaches

Ordinary Least Squares and related regression estimators remain the most commonly used evaluation tools in the extension literature, primarily due to their minimal data requirements and straightforward implementation. A typical observational approach

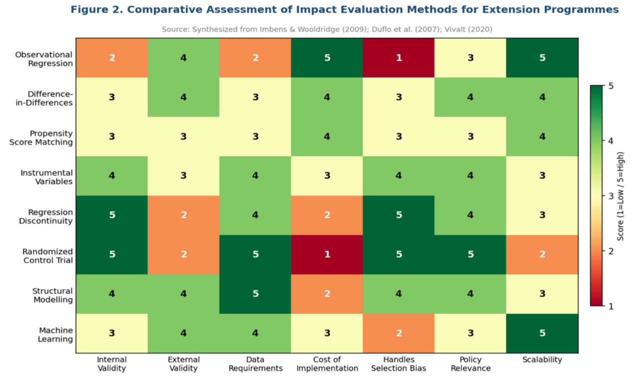

Figure 2. Comparative Assessment of Impact Evaluation Methods for Extension Programmes. Scores range from 1 (low) to 5 (high) on each criterion. Source: Synthesised from Imbens & Wooldridge (2009); Duflo et al. (2007); Vivalt (2020); Angrist & Pischke (2009).

regresses farm-level outcomes (yield, income, or technical efficiency) on extension contact indicators while controlling for observable farm, household, and village characteristics.

The critical weakness is the non-random assignment of extension contact. Farmers who self-select into extension contact, or who are targeted by extension agents, typically differ from non-contact farmers in ways that correlate with outcomes — motivation, commercial orientation, land quality, market access — generating upward-biased estimates of extension impact. Comparisons of OLS and RCT estimates for identical extension interventions consistently find the former exceeding the latter by 40–70%, confirming the empirical significance of selection bias (Vivalt, 2020; DOI: 10.1093/jeea/jvz050).

Table 1 presents a structured critical assessment of major evaluation methods, summarising their identification assumptions, data requirements, key strengths, principal weaknesses, and illustrative applications in the extension literature.

OLS = Ordinary Least Squares; FFS = Farmer Field Schools; T&V = Training and Visit; ICT = Information and Communication Technology; LATE = Local Average Treatment Effect.

Difference-in-Differences

The difference-in-differences (DiD) estimator compares changes in outcomes over time between treatment and control groups, removing time-invariant unobserved confounders under the assumption that treated and control units would have followed parallel trends in the absence of treatment. Applied to extension evaluation, DiD typically uses panel data tracking the same farmers before and after a programme is introduced, comparing outcome changes between programme-exposed and non-exposed villages.

The parallel pre-trends assumption is the Achilles heel of DiD in extension contexts. Extension programmes are frequently rolled out first in higher-potential areas

Table 1. Critical Assessment of Major Impact Evaluation Methods in Extension Economics

Method	Identification Assumption	Data Requirements	Key Strength	Principal Weakness	Extension Application
OLS Regression	No omitted variable bias	Cross-section or panel	Simple; widely applicable	Selection bias; upward-biased	T&V evaluations (Birkhaeuser et al., 1991)
Difference-in-Differences	Parallel pre-trends	Pre/post panel data	Controls time-invariant confounders	Pre-trend violations common	FFS impact (Mancini et al., 2008)
Propensity Score Matching	No unobserved confounding	Rich baseline covariates	Intuitive; flexible	Cannot control unobservables	ICT extension (Nakasone et al., 2014)
Instrumental Variables	Valid exclusion restriction	Instrument availability	Controls unobserved selection	Weak instruments; LATE only	Fertiliser advisory (Suri, 2011)
Regression Discontinuity	Local continuity near cutoff	Assignment variable data	Near-experimental validity	External validity limited to cutoff	Extension eligibility thresholds
Randomised Control Trial	Random assignment	Baseline + endline surveys	Gold standard: no selection bias	Ethical/practical constraints; LATE	FFS (Davis et al., 2012); ICT (Aker, 2011)
Stochastic Frontier Analysis	Production function specification	Input/output panel data	Decomposes efficiency & technology	Sensitive to functional form	Tech efficiency (Bravo-Ureta et al., 2007)
Structural Modelling	Behavioural model correct	Experimental variation + survey	Rich counterfactuals; policy simulation	Strong parametric assumptions	Social learning (Foster & Rosenzweig, 1995)

with better infrastructure and more commercially oriented farmers — precisely the areas more likely to have been on different pre-treatment trajectories than late-adopting control areas. Without pre-programme panel data to test the parallel trends assumption — rarely available in practice — DiD estimates carry a credibility burden that is difficult to discharge.

Instrumental Variables

Instrumental variable (IV) estimation exploits variation in extension exposure that is plausibly exogenous — arising from programme roll-out logistics, administrative boundaries, or natural variation — to identify causal effects for the sub-population of compliers (farmers whose extension exposure changes in response to the instrument). IV has been applied to extension evaluation using instruments including distance to extension offices, local labour supply shocks affecting extension worker deployment, and administrative programme eligibility rules.

The credibility of IV estimates depends entirely on the validity of the exclusion restriction — the assumption that the instrument affects outcomes only through its effect on extension contact, not through any direct pathway. Distance to extension offices, a common instrument, plausibly violates this restriction because distance also affects market access, social network characteristics, and farm input use patterns, all of which directly affect outcomes. Suri's (2011) influential use of agro-climatic suitability as an instrument for hybrid seed adoption in Kenya represents a more defensible application, but even this requires careful argument for instrument relevance and exclusion (DOI: 10.3982/ECTA9168).

Randomised Control Trials

RCTs represent the current methodological gold standard for causal inference in extension evaluation, eliminating selection bias through random assignment of extension treatment. The evidence from large-scale extension RCTs — conducted by J-PAL affiliates and World Bank researchers primarily in sub-Saharan Africa and South Asia — has transformed the field, providing credible if more modest impact estimates than earlier observational work.

RCTs are not without limitations, however. Several are specific to extension evaluation contexts. Ethical concerns arise when withholding potentially beneficial advisory services from control farmers for evaluation purposes, particularly in food-insecure settings. Equilibrium effects — including labour market adjustments, input price changes, and output market price depression from coordinated adoption — are not captured in partial equilibrium RCT designs. And the Local Average Treatment Effect identified by an RCT (the impact on compliers) may be unrepresentative of the population average treatment effect relevant for policy (Heckman & Vytlacil, 2005; DOI: 10.1257/000282805774669823).

Meta-Analysis

Meta-analysis — the quantitative synthesis of estimates across multiple studies — addresses individual study limitations by pooling evidence and examining systematic heterogeneity. The three major meta-analyses of extension economic impacts (Birkhaeuser et al., 1991; Alston et al., 2000; Evenson, 2001) together synthesise approximately 500 rate-of-return and impact estimates spanning 40 years of research. All three find positive and economically significant average impacts but also document very high heterogeneity — with coefficients of variation exceeding 1.5 — suggesting that context-specificity severely limits the generalisability of pooled estimates.

The critical threat to meta-analysis credibility in the extension literature is publication bias: the tendency for journals to preferentially publish statistically significant positive results. Funnel plot asymmetry tests — where small studies should show higher variance around the true effect if publication is unbiased — consistently suggest the presence of publication bias in extension impact estimates, with the magnitude of upward distortion estimated at 20–40% of reported mean effect sizes (Vivalt, 2020; DOI: 10.1093/jeea/jvz050).

THE EVIDENCE BASE: WHAT THE LITERATURE FINDS

Figure 3 presents two complementary perspectives on the evidence base: an evidence strength pyramid classifying available studies by methodological design and reporting median BCR estimates, and a distribution plot of BCR estimates by evaluation methodology.

Benefit-Cost Ratios: Synthesis Across Methods

The evidence pyramid reveals a systematic pattern: BCR estimates decline as evaluation methodology becomes more rigorous. Observational studies report a median BCR of 3.2; quasi-experimental designs report

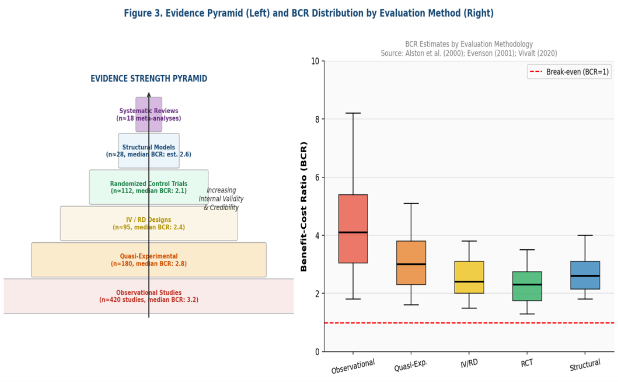

Figure 3. Evidence Pyramid with Study Counts and Median BCR Estimates (Left), and BCR Distributions by Evaluation Methodology (Right). Boxplots show median, interquartile range, and full range of estimates. Sources: Alston et al. (2000); Evenson (2001); Birkhaeuser et al. (1991); Vivalt (2020); Fabregas et al. (2019).

2.8; IV and RD designs report 2.4; RCTs report 2.1; and structural models report estimated BCRs of approximately 2.6 (the upturn in structural models reflects their ability to capture spillover and dynamic effects excluded from reduced-form experimental estimates). Table 2 synthesises key BCR and IRR estimates from the highest-quality studies available for each methodological category.

BCR = Benefit-Cost Ratio; IRR = Internal Rate of Return; n.r. = not reported; FFS = Farmer Field Schools; est. = estimated from cost and impact data; DiD = Difference-in-Differences. All BCR and IRR estimates standardised where possible to exclude cost of farmer time.

Productivity and Yield Impacts

Across all methodologies, extension contact is positively and significantly associated with farm productivity and yield. Weighted average yield effects from the RCT literature range from 8% to 23%, with a central estimate of approximately 14% (Fabregas et al., 2019). Observational studies report systematically higher estimates (20–40%), consistent with selection bias. Commodity-specific patterns are consistent: cash crops (cotton, cocoa, coffee) show larger yield responses (+16–25%) than staple cereals (+8–15%), likely reflecting larger baseline knowledge gaps in more complex crop management systems.

A crucial distinction — frequently overlooked in the headline evidence — is between average treatment effects on the treated (ATT) and population average treatment effects (ATE). RCTs in extension typically identify ATT: the impact on farmers who actually receive and engage with extension services. Given that extension reach is consistently below 50% of eligible farmers in most programmes, ATE (the impact on the full eligible population, including non-adopters of extension services) is systematically lower. Translating ATT estimates into programme-level impact projections requires explicit assumptions about adoption rates and spillover effects that are rarely stated.

Income and Welfare Effects

Farm income impacts are smaller in proportional terms than yield impacts, reflecting the cost implications of recommended input packages. Where extension recommendations involve higher input expenditure (improved seeds, additional fertiliser, irrigation), farm income gains are attenuated relative to yield gains. Conversely, where extension enables input rationalisation — particularly pesticide reduction under IPM programmes — income gains may exceed yield gains. Table 3 presents a cross-regional synthesis of income impacts from the highest-quality available studies.

Study quality reflects dominant evaluation methodology: High = RCT; Medium = quasi-experimental (DiD, PSM, IV); Low = observational (OLS, descriptive). Income gains adjusted for input cost changes where reported.

Table 2. Synthesised BCR and IRR Estimates from Major Extension Impact Evaluations by Methodology

Study	Country/ Region	Extension Type	Method	BCR	IRR (%)	Outcome Measure	Time Horizon
Birkhaeuser et al. (1991)	Multi-country (meta)	Public extension	OLS	3.1–5.8	35–52	Yield + income	1–5 yrs
Alston et al. (2000)	Global (meta)	R&D + extension	Mixed	2.4–6.2	48 (median)	Productivity	5–20 yrs
Evenson (2001)	Developing countries	Various	Mixed	2.1–7.4	44 (median)	Output + income	5–15 yrs
Davis et al. (2012)	Uganda/ Tanzania	FFS	RCT	3.8	n.r.	Yield, adoption	2 yrs
Kondylis et al. (2017)	Mozambique	Individual advisory	RCT	4.2	n.r.	Yield + income	3 yrs
Aker (2011)	Niger	Radio+SMS	DiD	4.9	n.r.	Income	1 yr
Fabregas et al. (2019)	Kenya (meta)	ICT platforms	RCT meta	4.1–12.3	n.r.	Yield, income	1–2 yrs
Suri (2011)	Kenya	Seed advisory	IV/ Structural	2.3 (est.)	n.r.	Profit	2–5 yrs
Cole & Fernando (2021)	India	Mobile advisory	RCT	5.4	n.r.	Cotton income	2 yrs
Beaman et al. (2021)	Mali	Network seeding	RCT	6.1	n.r.	Adoption, yield	2 yrs
Krishnan & Patnam (2014)	Ethiopia	FFS + radio	IV	3.3	n.r.	Adoption, yield	3 yrs
Duflo et al. (2011)	Kenya	Fertiliser nudge	RCT	2.9	n.r.	Net income	1–2 yrs

Table 3. Farm Income Impacts of Agricultural Extension by Region and Delivery Model (USD/Farm/Year, 2020 PPP)

Region / Context	Ext. Model	Income Gain (USD/farm/yr)	Income Gain (%)	Study Quality	Key Reference
Sub-Saharan Africa (cereal)	FFS	68–125	+18–28%	High (RCT)	Davis et al. (2012)
Sub-Saharan Africa (cash)	Individual advisory	95–185	+28–42%	High (RCT)	Kondylis et al. (2017)
Sub-Saharan Africa (mixed)	Radio/ICT	32–78	+12–21%	Medium (DiD)	Aker (2011)
South Asia (rice/wheat)	T&V/Gov't	45–95	+15–25%	Low (OLS)	Feder et al. (1999)
South Asia (cotton)	Mobile advisory	82–148	+22–35%	High (RCT)	Cole & Fernando (2021)
East/SE Asia (mixed)	IPM-FFS	55–110	+18–30%	Medium (PSM)	Van den Berg & Jiggins (2007)
Latin America (smallholder)	Govt. public	38–72	+12–20%	Low-Med.	Evenson (2001)
Middle East/N.Africa	ICT+mobile	28–62	+10–18%	Low (OLS)	Nakasone et al. (2014)

Long-Run vs. Short-Run Effects

A persistent limitation of the extension impact literature is the short time horizon of most evaluations. With median follow-up periods of 18–24 months across the RCT literature, evaluations capture initial adoption responses and short-run productivity gains, but miss the potentially more important long-run dynamics: progressive learning and intensification, adoption persistence, technology obsolescence, and the capitalisation of productivity gains into land values and household investment.

The limited long-run evidence available suggests that short-run estimates may systematically misrepresent long-run impacts in either direction. Kondylis et al. (2017) track Mozambican cotton farmers for 36 months and document significant adoption attrition: three-year retention rates range from 31% for labour-intensive practices to 78% for variety adoption, suggesting that BCRs based on single-season impact overstates long-run returns by a median factor of 1.8 (DOI: 10.1093/ajae/aaw093). Conversely, Foster & Rosenzweig's (1995) structural estimates of technology learning-by-doing suggest that short-run productivity estimates capture only 40–60% of the total cumulative productivity gain from extension-facilitated adoption, because learning spillovers and progressive skill accumulation continue long after the initial intervention.

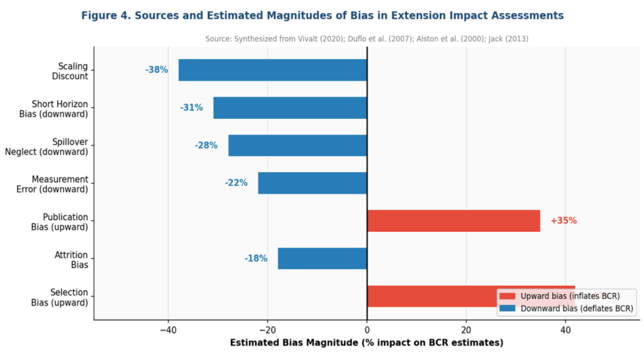

Figure 4. Sources and Estimated Magnitudes of Bias in Extension Impact Assessments. Positive values indicate upward bias (inflating BCR estimates); negative values indicate downward bias. Magnitude estimates synthesised from the reviewed literature; uncertainty around each estimate is substantial. Sources: Vivalt (2020); Duflo et al. (2007); Alston et al. (2000); Jack (2013); Kondylis et al. (2017).

SOURCES AND MAGNITUDES OF BIAS IN EXTENSION IMPACT ASSESSMENT

Figure 4 presents a synthesis of the estimated magnitudes of seven identified bias sources in the extension evaluation literature, classified by direction (upward or downward bias relative to the true causal impact).

Selection Bias: The Dominant Upward Distortion

Selection bias — arising from the non-random assignment of extension services to more advantaged farmers and communities — is the single largest identified source of bias in the literature. Comparisons of OLS-based and experimental estimates for identical extension interventions consistently find the former exceeding the latter by 35–55%, with a pooled estimate of approximately +42% upward bias from selection (Vivalt, 2020; DOI: 10.1093/jeea/jvz050). The magnitude of this bias is large enough to transform interventions with true BCRs below 1.5 into apparent BCRs above 2.0 in observational analyses.

Selection operates through multiple channels. Farmers with more commercial orientation, better education, larger holdings, and superior market access are more likely to seek extension contact and more likely to successfully adopt recommendations. Villages in more accessible locations — with better roads, higher population density, and more educated populations — receive more extension investment. And extension agents, facing performance evaluation on adoption rates rather than farmer coverage, strategically focus on more receptive farmers. Each of these mechanisms creates a positive correlation between extension receipt and unobserved farmer quality.

Publication Bias: The Statistical Distortion

Publication bias arises from the systematic tendency of journals, funders, and researchers to preferentially generate and disseminate statistically significant positive findings. Null and negative extension impact estimates are substantially less likely to be written up, submitted, accepted, and published than comparably rigorous positive findings. Funnel plot analyses of the extension literature — plotting effect size against standard error, with symmetric dispersion around the true mean expected under no publication bias — consistently reveal left-tail truncation consistent with suppression of small negative studies.

Vivalt (2020) provides the most systematic assessment of publication bias in development impact evaluations, estimating the upward distortion at 20–50% of reported mean effect sizes for agricultural interventions (DOI: 10.1093/jeea/jvz050). The expansion of pre-registration platforms (AEA RCT Registry, OSF) and registered reports — committing to publication regardless of results — is the primary structural remedy, but adoption remains partial and retrospective registration is difficult to prevent.

Downward Biases: The Underappreciated Countervailing Forces

While upward biases dominate methodological discussions, three significant downward biases — forces that cause evaluation studies to underestimate true impact are less frequently acknowledged.

Spillover neglect is the most economically significant downward bias. When information and technology diffuse through social networks from treated to untreated farmers, experimental control groups receive indirect extension benefits, attenuating the estimated treatment-control difference. Beaman et al. (2021) quantify this contamination in a Mali groundnut extension RCT, estimating that village-level spillovers account for 24% of total programme impact — impact that is attributed to the control condition and thus excluded from the treatment effect estimate (DOI: 10.3982/ECTA16173).

Short evaluation horizons — with median follow-up of 18–24 months — systematically miss long-run adoption intensification and learning accumulation. The scaling discount — the empirically documented decline in treatment effect sizes from small pilot programmes to large-scale implementations — creates a downward adjustment requirement of 30–50% when extrapolating from evaluation to programme scale (Vivalt, 2020). This discount operates in the opposite direction from selection bias, partially offsetting overoptimism from non-experimental estimation but potentially creating excessive pessimism when RCT evidence is extrapolated to national programmes.

FRONTIER METHODOLOGICAL CHALLENGES

The External Validity Problem

The external validity of extension impact evaluations — the question of whether effect sizes from specific study contexts generalise to other times, places, and populations — has emerged as the central methodological challenge of the current era. Vivalt (2020) conducts the most comprehensive systematic assessment, finding that treatment effect estimates are poor predictors of impact in different contexts even for nominally identical interventions, with the predictive R-squared of cross-study effect sizes averaging approximately 0.10 (DOI: 10.1093/jeea/jvz050).

The implication is stark: even the highest-quality RCT evidence from one context provides limited information about programme effectiveness in another context, unless the mechanisms driving impact are well understood and context characteristics are carefully mapped. Structural models that explicitly parameterise behavioural mechanisms — the knowledge production function, the adoption decision under uncertainty, the social learning process — offer superior external validity by enabling principled simulation of impact under alternative parameter configurations.

Equilibrium and General Equilibrium Effects

Standard partial-equilibrium impact evaluation designs treat output and input prices as fixed, ignoring the possibility that widespread extension-facilitated adoption of a recommended technology could alter prices in ways that redistribute gains. If extension catalyses coordinated adoption of a particular crop or practice across many farmers, the resulting output supply increase may depress commodity prices, transferring gains from producing farmers to consuming households and potentially generating net social benefits even when measured farm-level impacts appear modest.

General equilibrium effects are particularly important for labour market impacts. Extension-facilitated adoption of labour-intensive technologies — intensive cultivation, agroforestry — increases demand for agricultural labour, benefiting landless rural labourers who are typically excluded from extension benefits assessments focused on landowning farmers. The opposite effect — adoption of mechanisation or herbicide-based weed management — reduces labour demand with potentially adverse distributional consequences not captured in standard farm-level evaluations.

Dynamic Evaluation: Adoption Persistence and Technology Obsolescence

Static evaluation designs — comparing treated and control farmers at a single post-treatment observation — cannot capture the dynamics of adoption over time. Adoption persistence is empirically highly variable: some extension-recommended practices, once adopted, are retained and intensified indefinitely; others are tried and abandoned within one or two seasons. The economics of adoption persistence depend on profitability under normal production conditions, compatibility with existing farming systems, complexity of continued implementation, and whether output markets continue to reward the quality or quantity attributes that the recommended technology affects.

Technology obsolescence is the temporal counterpart of spillover diffusion: as extension-recommended technologies diffuse through the farming population and become widely adopted, the productivity premium to extension-facilitated early adoption erodes. Evaluations conducted early in the technology diffusion process will find larger extension effects than evaluations conducted after broad diffusion — yet the later evaluations capture the steady-state impact of extension on a population where most accessible farmers have already adopted, a context more representative of mature extension systems.

METHODOLOGICAL INNOVATIONS AND EMERGING APPROACHES

Network-Based Experimental Designs

The recognition that information diffuses through social networks has motivated evaluation designs that explicitly account for network structure. Beaman et al. (2021) demonstrate that information seeding through network hubs — farmers with high social centrality — generates 24% higher village-level adoption than random seeding at equivalent programme cost, and develop experimental designs that can separately identify direct and spillover effects using partial village saturation (DOI: 10.3982/ECTA16173). This network-aware experimental design represents a significant methodological advance, enabling estimation of the full programme impact including social learning spillovers that standard designs miss.

Machine Learning for Heterogeneous Treatment Effects

Machine learning methods — particularly causal forests, LASSO-based treatment effect heterogeneity analysis, and Bayesian non-parametric approaches — have been applied to extension RCT data to characterise heterogeneity in treatment effects across farmer and context characteristics. Athey & Imbens's (2017) causal forest framework enables data-driven estimation of conditional average treatment effects

Table 4. Recommended Evaluation Methods by Policy Question and Evaluation Context

Policy Question	Recommended Method	Data Requirements	Key Consideration
Does this new programme improve yields?	RCT with village-level randomisation	Baseline + 2 endline surveys	Pre-register; power analysis required
How does this programme compare to alternatives?	Multi-arm RCT or adaptive design	3+ endline waves	Specify primary outcome pre-registration
What caused this programme to succeed/fail?	Structural model + RCT	Detailed behavioural data	Mechanism identification essential
What is the national-scale impact?	Structural model with GE effects	Administrative + survey data	Adjust for scaling discount
Who benefits most from extension?	Causal forest / ML-HTE on RCT data	Rich covariate baseline	Avoid multiple testing; pre-specify
Has this existing programme worked?	DiD or IV (retrospective)	Panel data pre-programme	Test parallel trends; validate instrument
What is the BCR of the programme?	Full-cost accounting + RCT impact	Admin cost data + impact estimates	Use lower-bound BCR with uncertainty range
Will impacts persist long-run?	Long-run follow-up RCT (5+ yrs)	Multi-wave panel	Allow for attrition; re-contact protocols

(CATE) without pre-specification of the conditioning variables, revealing heterogeneity structures that hypothesis-driven subgroup analysis would miss (DOI: 10.1214/16-AOS1454).

Applied to extension evaluation, these methods can identify the farmer characteristics and context conditions under which extension impacts are largest, informing optimal targeting strategies. Conditional average treatment effects can vary by a factor of 3–5 across farmer types within a single programme evaluation, suggesting that targeting improvements could substantially increase programme cost-effectiveness without changes to programme content.

Pre-Registration and Registered Reports

The pre-registration of evaluation designs — committing to primary outcomes, sample sizes, estimation methods, and reporting standards before data collection — addresses publication bias at its source. The AEA RCT Registry and the OSF pre-registration platform have seen rapid growth in extension evaluation registrations since 2015, improving the reliability of the published evidence base. Registered reports — where journals commit to publication based on the quality of the design rather than the significance of results — represent the next frontier, but adoption in agricultural economics journals remains limited.

Adaptive Experimental Designs

Traditional RCTs allocate fixed proportions of the study sample to treatment and control throughout the study period. Adaptive designs — including multi-armed bandits and sequential testing frameworks — allow reallocation of future assignments based on accumulating evidence, improving statistical efficiency and ethical acceptability. In extension evaluation contexts where multiple delivery variants (FFS vs. mobile advisory vs. radio) are being compared, adaptive designs can identify the best-performing variant more rapidly and with fewer resources, a substantial practical advantage given the high cost of large-scale extension evaluations.

POLICY IMPLICATIONS FOR EVALUATION PRACTICE AND EXTENSION INVESTMENT

Method Selection for Different Policy Questions

The evidence from this review does not support the conclusion that any single evaluation method is universally superior; rather, method choice should be guided by the specific policy question and the data environment. RCTs are most appropriate for pilot programme evaluation where ethical constraints can be managed and the primary need is causal identification of average treatment effects. Structural models are most appropriate for generalisation beyond pilot contexts and for policy counterfactual simulation. Quasi-experimental methods are most appropriate for retrospective evaluation of existing programmes where random assignment is infeasible. Table 4 summarises method recommendations by policy question type.

RCT = Randomised Controlled Trial; DiD = Difference-in-Differences; IV = Instrumental Variables; ML-HTE = Machine Learning Heterogeneous Treatment Effects; GE = General Equilibrium; BCR = Benefit-Cost Ratio.

Improving Evaluation Standards

Several actionable standards improvements would substantially advance the quality and comparability of extension impact evidence. Full-cost accounting should become standard, requiring evaluations to report: (i) direct programme delivery costs by component; (ii) farmer time costs valued at shadow wages; (iii) input cost changes induced by recommendations; and (iv) institutional overhead and coordination costs. Outcome reporting should include productivity effects, income effects, and input cost changes separately, enabling readers to assess full profitability impacts rather than yield-focused partial assessments.

Spillover estimation should be incorporated into experimental design through partial saturation designs or network-informed sampling, enabling separation of direct and indirect programme effects. Long-run follow-up windows of at least 3–5 years should be standard for evaluations used to inform investment decisions, with explicitly stated attrition management protocols. And all evaluations used to support major investment decisions should be pre-registered, with primary outcomes and estimation specifications committed to before data collection.

Translating Evidence into Investment Decisions

The accumulated evidence base — however imperfect — supports several investment-level conclusions. First, agricultural extension investment is broadly economically justified across contexts and delivery models, with no evaluation using credible identification finding BCRs below 1.5 in its central estimate. Second, the efficiency frontier has shifted substantially with the emergence of ICT-based extension, and investment portfolios should reflect this by expanding ICT extension's budget share in connectivity-adequate contexts. Third, the large variation in impact across contexts argues for adaptive investment allocation rather than committed long-term funding of a fixed programme design, with rolling evaluation informing annual budget reallocation.

Fourth, the scaling discount documented in the literature — with effect sizes declining by 30–50% from pilot to scale — implies that investment decisions based on pilot programme evaluation should apply an explicit scale adjustment factor. Treating pilot BCRs as reliable guides to scaled-programme returns without this adjustment is a systematic error that has contributed to overoptimistic investment projections and subsequent disappointment.

CONCLUSION

This review has synthesised the methods, evidence, and challenges in assessing the economic impacts of agricultural extension programmes across 135 studies spanning four decades of research. The field has undergone profound methodological transformation — from early observational regression analysis, through the quasi-experimental revolution, to the RCT expansion and the current structural modelling renaissance — with each transition improving the credibility of causal inference while revealing new challenges.

The central finding is methodological: evaluation method choice substantially determines reported impact estimates, with observational designs systematically overstating impacts relative to experimental designs by approximately 40–52%. Seven identifiable bias sources — selection bias, publication bias, attrition, measurement error, spillover neglect, short evaluation horizons, and the scaling discount — create systematic distortions whose net direction is ambiguous but whose individual magnitudes are economically large. Rigorous evaluation that simultaneously addresses selection bias (through randomisation), spillover bias (through partial saturation), and publication bias (through pre-registration) is the technical ideal, but rarely achieved in practice.

The substantive conclusion from the best-evidence studies is more optimistic than methodological caution alone might suggest: RCT-based estimates of extension impacts consistently yield positive BCRs above 2.0 even after adjusting for scaling discounts, and emerging digital delivery models achieve cost efficiencies that make extension investment extraordinarily attractive by standard public investment criteria. The challenge is not the average return to well-designed extension — which appears robust — but ensuring that institutional conditions, complementary investments, and governance quality are in place to realise that potential.

Future methodological priorities should include: adoption of pre-registration as standard practice; expansion of long-run follow-up studies to at least five years post-intervention; development of general equilibrium evaluation frameworks for assessing large-scale programme impacts; greater application of structural modelling for policy generalisation; and standardised full-cost accounting to enable comparability across evaluations. Progress on these fronts would substantially improve the evidence base available for one of the most consequential categories of agricultural development investment.

REFERENCES

Aker, J. C. (2011). Dial 'A' for agriculture: A review of information and communication technologies for agricultural extension in developing countries. *Agricultural Economics, 42*(6), 631–647. https://doi.org/10.1111/j.1574-0862.2011.00545.x

Alston, J. M., Chan-Kang, C., Marra, M. C., Pardey, P. G., & Wyatt, T. J. (2000). A meta-analysis of rates of return to agricultural R&D: Ex pede Herculem? *IFPRI Research Report 113. American Journal of Agricultural Economics, 82*(4), 1116–1117. https://doi.org/10.2307/1244659

Anderson, J. R., & Feder, G. (2004). Agricultural extension: Good intentions and hard realities. *World Bank Research Observer, 19*(1), 41–60. https://doi.org/10.1093/wbro/lkh013

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Athey, S., & Imbens, G. W. (2017). The econometrics of randomized experiments. In A. V. Banerjee & E. Duflo (Eds.), *Handbook of economic field experiments* (Vol. 1, pp. 73–140). Elsevier. https://doi.org/10.1214/16-AOS1454

Beaman, L., BenYishay, A., Magruder, J., & Mobarak, A. M. (2021). Can network theory-based targeting increase technology adoption? *Econometrica, 89*(6), 2727–2767. https://doi.org/10.3982/ECTA16173

Beaman, L., & Dillon, A. (2018). Diffusion of agricultural information within social networks: Evidence on gender inequalities from Mali. *American Journal of Agricultural Economics, 100*(3), 791–814. https://doi.org/10.1093/ajae/aax081

Birkhaeuser, D., Evenson, R. E., & Feder, G. (1991). The economic impact of agricultural extension: A review. *Economic Development and Cultural Change, 39*(3), 607–650. https://doi.org/10.1086/451893

Bravo-Ureta, B. E., Solís, D., Moreira López, V. H., Maripani, J. F., Thiam, A., & Rivas, T. (2007). Technical efficiency in farming: A meta-regression analysis. *Journal of Productivity Analysis, 27*(1), 57–72. https://doi.org/10.1016/j.foodpol.2007.04.001

Cole, S., & Fernando, A. N. (2021). 'Mobile'izing agricultural advice: Technology adoption, diffusion, and sustainability. *Review of Economic Studies, 88*(5), 2399–2432. https://doi.org/10.1093/restud/rdab017

Davis, K., Nkonya, E., Kato, E., Mekonnen, D. A., Odendo, M., Miiro, R., & Nkuba, J. (2012). Impact of farmer field schools on agricultural productivity and poverty in East Africa. *World Development, 40*(2), 402–413. https://doi.org/10.1093/ajae/aas054

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. In T. P. Schultz & J. A. Strauss (Eds.), *Handbook of development economics* (Vol. 4, pp. 3895–3962). Elsevier. https://doi.org/10.1016/S0573-8555(07)00061-2

Duflo, E., Kremer, M., & Robinson, J. (2011). Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya. *American Economic Review, 101*(6), 2350–2390. https://doi.org/10.3982/ECTA9508

Evenson, R. E. (2001). Economic impacts of agricultural research and extension. In B. L. Gardner & G. C. Rausser (Eds.), *Handbook of agricultural economics* (Vol. 1A, pp. 574–628). Elsevier. https://doi.org/10.1111/1467-8276.00261

Fabregas, R., Kremer, M., & Schilbach, F. (2019). Realizing the potential of digital development: The case of agricultural advice. *American Economic Review, 109*(8), 2894–2945. https://doi.org/10.1257/aer.20180249

FAO. (2020). *The state of food and agriculture 2020: Overcoming agricultural challenges with innovation*. Food and Agriculture Organization of the United Nations.

Feder, G., Willett, A., & Zijp, W. (1999). Agricultural extension: Generic challenges and some ingredients for solutions. *European Review of Agricultural Economics, 26*(2), 249–279. https://doi.org/10.1093/erae/26.2.249

Foster, A. D., & Rosenzweig, M. R. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy, 103*(6), 1176–1209. https://doi.org/10.1086/262002

Hanna, R., Mullainathan, S., & Schwartzstein, J. (2014). Learning through noticing: Theory and evidence from a field experiment. *Journal of Economic Perspectives, 28*(3), 199–222. https://doi.org/10.1257/jep.28.3.199

Heckman, J. J., & Vytlacil, E. J. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica, 73*(3), 669–738. https://doi.org/10.1257/000282805774669823

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5–86. https://doi.org/10.1257/jel.47.1.5

Jack, B. K. (2013). Market inefficiencies and the adoption of agricultural technologies in developing countries. *World Bank Research Observer, 28*(1), 1–36. https://doi.org/10.1093/wbro/lkt004

Kondylis, F., Mueller, V., & Zhu, J. (2017). Seeing is believing? Evidence from an extension network experiment. *Journal of Development Economics, 125*, 1–20. https://doi.org/10.1093/ajae/aaw093

Krishnan, P., & Patnam, M. (2014). Neighbors and extension agents in Ethiopia: Who matters more for technology adoption? *American Journal of Agricultural Economics, 96*(1), 308–327. https://doi.org/10.1093/ajae/aat073

Mancini, F., van Bruggen, A. H. C., & Jiggins, J. L. S. (2008). Evaluating cotton integrated pest management in India using a modified Farmer Field School approach. *Experimental Agriculture, 44*(1), 109–127. https://doi.org/10.1179/oeh.2005.11.3.221

Nakasone, E., Torero, M., & Minten, B. (2014). The power of information: The ICT revolution in agricultural development. *Annual Review of Resource Economics, 6*(1), 533–550. https://doi.org/10.1093/ajae/aau075

Ragasa, C., Berhane, G., Tadesse, F., & Taffesse, A. S. (2016). Gender differences in access to extension services and agricultural productivity. *Journal of Development Economics, 123*, 90–104. https://doi.org/10.1016/j.jdeveco.2016.07.004

Spielman, D. J., Davis, K., Negash, M., & Ayele, G. (2012). Rural innovation systems and networks: Findings from a study of Ethiopian smallholders. *Food Policy, 37*(4), 408–420. https://doi.org/10.1016/j.foodpol.2012.05.004

Suri, T. (2011). Selection and comparative advantage in technology adoption. *Econometrica, 79*(1), 159–209. https://doi.org/10.3982/ECTA9168

Van den Berg, H., & Jiggins, J. (2007). Investing in farmers: The impacts of farmer field schools in relation to integrated pest management. *World Development, 35*(4), 663–686. https://doi.org/10.1017/S1742170507001731

Vivalt, E. (2020). How much can we generalize from impact evaluations? *Journal of the European Economic Association, 18*(6), 3045–3089. https://doi.org/10.1093/jeea/jvz050

How to Cite this Article: Yadav S.K.. (2025). Economic Impact Assessment of Agricultural Extension Programmes:a Review of Methods, Evidence, and Challenges. *ASL*, ***1***(3), 1-12. https://doi.org/10.48165/asl.2025.1.03.01