Does Access to Agricultural Credit Improve Smallholder Farmers' Productivity in Kenya? Evidence from Econometric Analysis

Danton Omondi, Dr Yasin Kuso Ghabon

Department of Economics, Maseno University

Corresponding author

DOI: https://doi.org/10.47772/IJRISS.2026.101100084

Received: 29 October 2026; Accepted: 18 November 2026; Published: 21 November 2026

ABSTRACT

Treating agricultural credit as exogenous overstates its effect on smallholder productivity: in an illustrative sample of 400 Kenyan-style farm households, OLS shows credit raising yields by about 57.9%, but two-stage least squares (2SLS) cuts the estimate to about 13.6%, which is not statistically significant. Policy debate in Kenya often assumes that extending credit to smallholders will raise farm output, yet borrowers differ from non-borrowers in unobserved managerial ability, which biases naive comparisons. This paper specifies a Cobb-Douglas log-yield model, derives the OLS and 2SLS estimators in matrix form, and works every formula through numerically. Distance to the nearest lender and membership of a farmer group or SACCO serve as instruments. The first-stage F-statistic is 45.76, the Sargan test does not reject instrument validity ($p = 0.488$), and the Durbin-Wu-Hausman test rejects the exogeneity of credit ($p = 0.008$). The results show why identification, not just regression, matters for evaluating rural credit policy. The data are simulated to mimic a household survey, so the findings illustrate the method and are not evidence about Kenya.

Keywords: agricultural credit; smallholder productivity; endogeneity; instrumental variables; two-stage least squares; Kenya

1.0 INTRODUCTION

1.1 Background

Agriculture remains central to Kenya's economy and to rural livelihoods, and most farm output comes from smallholders working a few acres each (Government of Kenya, 2019). Productivity on these farms depends on timely purchase of seed, fertiliser and hired labour, and these purchases require cash before harvest. Formal and semi-formal credit is the main way to bridge that gap. National policy therefore treats financial access as a lever for agricultural transformation, as seen in the Agricultural Sector Transformation and Growth Strategy and the Bottom-Up Economic Transformation Agenda.

The theoretical case is well known. Where information is imperfect, lenders ration credit and many productive farmers are shut out (Stiglitz and Weiss, 1981; Hoff and Stiglitz, 1990). Relaxing the constraint should raise input use and output (Feder et al., 1990; Conning and Udry, 2007). The empirical case is harder, because the farmers who obtain credit are rarely a random draw.

1.2 Problem statement

A simple comparison of borrowers and non-borrowers will usually show higher yields among borrowers. That gap need not be caused by credit. Farmers with greater managerial ability, better networks or stronger land rights are both more likely to be approved and more likely to farm well. If this unobserved ability sits in the error term, ordinary least squares (OLS) overstates the benefit of credit, and policy built on it may overspend on credit programmes that deliver less than promised. Studies that treat credit as exogenous therefore cannot answer the policy question: what happens to productivity when a farmer who would otherwise be excluded gains access?

1.3 Objectives

The general objective is to estimate the effect of access to agricultural credit on smallholder productivity. The specific objectives are to:

1. estimate the association between credit access and log yield per acre using OLS;
2. test whether credit access is endogenous in the yield equation;
3. estimate the causal effect of credit access on yield using instrumental variables (2SLS); and
4. test the validity and strength of the instruments and the classical assumptions of the model.

1.4 Hypotheses

The null hypotheses, each stated against a two-sided alternative, are:

- H_0^1: Credit access has no effect on smallholder yield, $\beta_1 = 0$.
- H_0^2: Credit access is exogenous in the yield equation, $\mathrm{Cov}(\mathrm{CREDIT}, u) = 0$.
- H_0^3: The excluded instruments are jointly irrelevant in the credit equation, $\pi_8 = \pi_9 = 0$.
- H_0^4: The overidentifying restrictions are valid, $\mathrm{Cov}(Z, u) = 0$.

1.5 Significance

The paper contributes a transparent, step-by-step treatment of identification in a rural credit setting. It is intended for students and analysts who need to see how each estimator is built from the data matrix, and it shows how the choice between OLS and 2SLS can change the policy conclusion. It also lays out a template that can be re-run on a real survey such as the Tegemeo panel or the Kenya Integrated Household Budget Survey.

2.0 LITERATURE REVIEW

2.1 Theoretical review

Credit rationing under imperfect information. Stiglitz and Weiss (1981) show that when lenders cannot observe borrower risk, a higher interest rate attracts riskier borrowers, so lenders ration credit instead of raising the price. Hoff and Stiglitz (1990) extend this to rural markets, where weak collateral, covariate shocks and costly monitoring make rationing common. Rationed smallholders cannot reach their optimal input mix, and their marginal products stay above the cost of capital.

Agricultural household and production-function approach. Feder et al. (1990) model how credit constraints reduce productivity when inputs cannot be financed. In the production-function framing used here, credit enters the log-linear Cobb-Douglas yield equation either as a shifter of the intercept or, indirectly, through higher input use. Because input use is controlled for directly, the credit coefficient captures any remaining effect on efficiency and technology choice.

Selection and endogeneity. Credit access is the outcome of two decisions: the farmer's decision to apply and the lender's decision to approve. Both depend on characteristics that the econometrician does not see. Wooldridge (2010) and Greene (2018) show that when a regressor is correlated with the error term, OLS is inconsistent and instruments are needed.

2.2 Empirical review

Evidence on credit and farm performance is mixed. Carter (1989) studied how credit constraints affected peasant productivity in Nicaragua and showed that effects differ with farm size. Feder et al. (1990) found a positive link between credit and productivity in Chinese agriculture using a disequilibrium model. Karlan et al. (2014), in a field experiment in northern Ghana, found that uninsured risk was a major constraint on farmers' investment alongside limited liquidity, which implies that credit alone may not lift productivity. Petrick (2005) reviews methods for identifying credit rationing and stresses that identification depends on how constrained farmers are classified.

For Kenya, the literature on credit access has mostly looked at who obtains credit and at adoption of improved practices. Feder et al. (1985) is the standard survey of adoption and shows that credit often matters through enabling the purchase of technology. Productivity studies that rely on OLS or propensity-score matching handle selection on observables but not on unobserved ability, which is why instruments or experiments are preferred.

2.3 Research gap

Three gaps motivate this paper. First, productivity regressions on credit often omit a formal test of endogeneity, so readers cannot tell whether OLS or IV estimates are appropriate. Second, studies that use instruments seldom report first-stage strength and overidentification tests alongside the estimates. Third, teaching material on this problem rarely shows the matrix algebra and arithmetic behind each estimator. This paper addresses the three gaps within one worked framework.

3.0 RESEARCH METHODOLOGY

3.1 Research design and data

The study uses a cross-sectional, quantitative design for household-level production data. The sample is $n = 400$ smallholder households, generated by computer simulation (random-number seed 805) to mimic the structure of a household survey. The data-generating process includes an unobserved managerial-ability term that raises both the chance of obtaining credit and the yield, and a true credit coefficient of 0.18 on log yield. The true value is known by construction, so each estimator can be judged against it. Because the data are simulated, results demonstrate the method; to answer the question for Kenya, the same code and tables should be re-run on survey data such as the Tegemeo panel or the Kenya Integrated Household Budget Survey.

Variable	Definition	Role
lnY	Log of crop yield (kg per acre)	Dependent
CREDIT	1 if the household obtained credit in the season, 0 otherwise	Endogenous regressor
lnLAND	Log of cultivated land (acres)	Control
lnLAB	Log of labour (man-days per acre)	Control
lnFERT	Log of fertiliser (kg per acre)	Control
EDUC	Years of schooling of the household head	Control
EXT	1 if the household had extension contact	Control
FEM	1 if the household head is female	Control
AGE	Age of the household head (years)	Control
DIST	Distance to the nearest lender (km)	Instrument
GROUP	1 if a member of a farmer group or SACCO	Instrument

3.2 Empirical model

A Cobb-Douglas technology is linear in logarithms. Credit and household characteristics shift the intercept,

giving the estimating equation:

$$\ln Y_i = \beta_0 + \beta_1\,\mathrm{CREDIT}_i + \beta_2 \ln \mathrm{LAND}_i + \beta_3 \ln \mathrm{LAB}_i + \beta_4 \ln \mathrm{FERT}_i + \beta_5\,\mathrm{EDUC}_i + \beta_6\,\mathrm{EXT}_i + \beta_7\,\mathrm{FEM}_i + \beta_8\,\mathrm{AGE}_i + u_i$$

Because $\ln Y$ is in logs and CREDIT is a dummy, the effect of credit on yield, expressed as a percentage, is:

$$\%\Delta Y = \left(e^{\beta_1} - 1\right) \times 100$$

3.3 OLS estimator in matrix form

Stack the n observations as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where $\mathbf{X}$ is $n \times k$ with $k = 9$ columns (a constant and eight regressors). The OLS estimator minimises the residual sum of squares $\mathbf{e}'\mathbf{e}$:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

With residuals $\mathbf{e} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$, the error variance and the covariance matrix of the estimator are:

$$\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k}, \qquad \widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}) = \hat{\sigma}^2\,(\mathbf{X}'\mathbf{X})^{-1}$$

The t-statistic for coefficient j is:

$$t_j = \frac{\hat{\beta}_j}{\mathrm{se}(\hat{\beta}_j)}, \qquad \mathrm{se}(\hat{\beta}_j) = \sqrt{\left[\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}})\right]_{jj}}$$

The fit and joint-significance measures, with $RSS = \mathbf{e}'\mathbf{e}$ and $TSS = \sum_i (y_i - \bar{y})^2$, are:

$$R^2 = 1 - \frac{RSS}{TSS}, \qquad \bar{R}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)}, \qquad F = \frac{(TSS - RSS)/(k-1)}{RSS/(n-k)}$$

In the one-regressor case used for the hand calculation, the matrix formula reduces to:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_i (x_i - \bar{x})\,(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

OLS is consistent only if $\mathrm{Cov}(\mathrm{CREDIT}, u) = 0$. That condition is doubtful here, because unobserved ability affects both credit and yield.

3.4 Instrumental variables and 2SLS

Let $\mathbf{Z}$ collect the exogenous regressors and the two excluded instruments, DIST and GROUP. The first stage regresses the endogenous variable on $\mathbf{Z}$:

$$\mathrm{CREDIT}_i = \pi_0 + \pi_1 \ln \mathrm{LAND}_i + \cdots + \pi_7\,\mathrm{AGE}_i + \pi_8\,\mathrm{DIST}_i + \pi_9\,\mathrm{GROUP}_i + v_i$$

The fitted values replace CREDIT in $\mathbf{X}$ to give $\widehat{\mathbf{X}} = \mathbf{P}_Z\mathbf{X}$, where $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is the projection matrix. The 2SLS estimator and its variance are:

$$\widehat{\boldsymbol{\beta}}_{2SLS} = \left(\widehat{\mathbf{X}}'\mathbf{X}\right)^{-1}\widehat{\mathbf{X}}'\mathbf{y} = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{y}$$

$$\widehat{\mathrm{Var}}\left(\hat{\boldsymbol{\beta}}_{2SLS}\right) = \hat{\sigma}_{IV}^2 \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}, \qquad \hat{\sigma}_{IV}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k}$$

Here $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{2SLS}$ is computed with the actual CREDIT, not its fitted values. For the instruments to be valid, they must be relevant ($\pi_8, \pi_9 \neq 0$) and excluded (uncorrelated with u). DIST affects the cost of reaching a lender, and GROUP gives access to group lending and loan guarantees; the exclusion restriction is that neither affects yield except through credit.

3.5 Specification tests and diagnostics

Instrument strength. The joint F-test on the two excluded instruments compares the restricted and unrestricted first-stage regressions. A value above 10 is the conventional rule of thumb against weak instruments (Staiger and Stock, 1997):

$$F_1 = \frac{(RSS_R - RSS_U)/q}{RSS_U/(n-k_Z)}, \qquad q = 2, \quad k_Z = 10$$

Endogeneity (Durbin-Wu-Hausman). Add the first-stage residual $\hat{v}$ to the OLS equation and test $\rho = 0$ with a t-test (Wu, 1973; Hausman, 1978). Rejection means CREDIT is endogenous and OLS is inconsistent:

$$\ln Y_i = \mathbf{x}_i'\boldsymbol{\beta} + \rho\,\hat{v}_i + \varepsilon_i, \qquad t_\rho = \frac{\hat{\rho}}{\mathrm{se}(\hat{\rho})}$$

Overidentification (Sargan). Regress $\hat{\mathbf{u}}$ on $\mathbf{Z}$ and compute the statistic below, which is distributed chi-square with (number of instruments $-$ number of endogenous regressors) $= 1$ degree of freedom under the null of valid instruments (Sargan, 1958):

$$S = nR_{\hat{u}}^2 \sim \chi^2(1)$$

Classical assumptions. Multicollinearity is checked with the variance inflation factor, where R_j^2 is from regressing regressor j on the others:

$$VIF_j = \frac{1}{1-R_j^2}$$

Heteroskedasticity is tested with Breusch-Pagan (1979): regress e_i^2 on $\mathbf{X}$ and compute

$$LM = nR_{aux}^2 \sim \chi^2(k-1)$$

Normality of residuals uses Jarque-Bera (1987), with S the residual skewness and K the kurtosis:

$$JB = \frac{n}{6}\left[S^2 + \frac{(K-3)^2}{4}\right] \sim \chi^2(2)$$

Functional form uses Ramsey's (1969) RESET, adding $\hat{y}^2$ and $\hat{y}^3$ to the model and testing them with an F-test. As a safeguard against heteroskedasticity, White-type (HC1) robust standard errors are also reported (White, 1980):

$$\widehat{\mathrm{Var}}_{HC1}\left(\hat{\boldsymbol{\beta}}\right) = \frac{n}{n-k}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\sum_{i=1}^{n} e_i^2\,\mathbf{x}_i\mathbf{x}_i'\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

4.0 RESULTS AND DISCUSSION

4.1 Descriptive statistics

Credit users are 43.25% of the sample ($173/400 = 0.4325$), and their mean yield is 1,057.75 kg per acre against 625.01 kg for non-users.

Variable	Mean	Std. dev.	Min	Max
Yield (kg per acre)	812.17	587.67	87.93	5,790.60
ln yield	6.4958	0.6383	4.4765	8.664
CREDIT (0/1)	0.4325	0.496	0	1
Land (acres)	1.7963	1.0701	0.3744	8
Labour (man-days per acre)	53.96	27.29	14.39	224.21
Fertiliser (kg per acre)	29.96	22.01	2.16	161.62
Education (years)	8.29	3.33	0	16
Extension contact (0/1)	0.3575	0.4799	0	1
Female head (0/1)	0.42	0.4942	0	1
Age (years)	43.83	11.45	22	73
DIST (km)	8.675	5.39	0.5	28.4
GROUP (0/1)	0.475	0.5	0	1

The raw gap in mean log yield between the two groups, with $n_1 = 173$ credit users and $n_0 = 227$ non-users, is:

$$\bar{y}_1 - \bar{y}_0 = 6.7843 - 6.2759 = 0.5084$$

A Welch t-test for the difference in means uses unequal variances:

$$t = \frac{\bar{y}_1 - \bar{y}_0}{\sqrt{s_1^2/n_1 + s_0^2/n_0}} = \frac{0.5084}{0.0594} = 8.55$$

The gap is highly significant, but it is a raw comparison with no controls and no correction for selection.

4.2 A worked example by hand ($n = 8$)

Before the full model, the estimator is computed by hand on eight farmers, with credit as the only regressor. The values are illustrative.

Farmer	1	2	3	4	5	6	7	8
CREDIT (x)	0	0	0	0	1	1	1	1
ln yield (y)	5.9	5.7	6.0	6.1	6.2	6.3	6.1	6.6

Step 1: sums and means. The sums are $\sum x_i = 4$, $\sum y_i = 48.9$, $\sum x_i^2 = 4$,

$\sum x_i y_i = 6.2 + 6.3 + 6.1 + 6.6 = 25.2$ and $\sum y_i^2 = 299.41$. The means are:

$$\bar{x} = \frac{4}{8} = 0.5, \qquad \bar{y} = \frac{48.9}{8} = 6.1125$$

Step 2: sums of squares and cross-products.

$$S_{xx} = \sum x_i^2 - n\bar{x}^2 = 4 - 8(0.5)^2 = 2.00$$

$$S_{xy} = \sum x_i y_i - n\bar{x}\bar{y} = 25.2 - 8(0.5)(6.1125) = 25.2 - 24.45 = 0.75$$

$$S_{yy} = \sum y_i^2 - n\bar{y}^2 = 299.41 - 8(6.1125)^2 = 299.41 - 298.90125 = 0.50875$$

Step 3: coefficients.

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{0.75}{2.00} = 0.375, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 6.1125 - 0.375(0.5) = 5.925$$

The intercept 5.925 equals the mean log yield of non-borrowers, and $5.925 + 0.375 = 6.300$ equals the mean of borrowers, as it must with a single dummy.

Step 4: the same result by matrix algebra. Here

$$\mathbf{X'X} = \begin{bmatrix} 8 & 4 \\ 4 & 4 \end{bmatrix}, \qquad \det(\mathbf{X'X}) = 8(4) - 4(4) = 16, \qquad \mathbf{X'y} = \begin{bmatrix} 48.9 \\ 25.2 \end{bmatrix}$$

$$(\mathbf{X'X})^{-1} = \frac{1}{16}\begin{bmatrix} 4 & -4 \\ -4 & 8 \end{bmatrix} = \begin{bmatrix} 0.25 & -0.25 \\ -0.25 & 0.50 \end{bmatrix}$$

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y} = \begin{bmatrix} 0.25(48.9) - 0.25(25.2) \\ -0.25(48.9) + 0.50(25.2) \end{bmatrix} = \begin{bmatrix} 12.225 - 6.300 \\ -12.225 + 12.600 \end{bmatrix} = \begin{bmatrix} 5.925 \\ 0.375 \end{bmatrix}$$

Step 5: residuals and inference. The fitted values are 5.925 for non-borrowers and 6.300 for borrowers, giving residuals of $-0.025, -0.225, 0.075, 0.175, -0.1, 0, -0.2, 0.3$. The residual sum of squares is:

$$RSS = \sum e_i^2 = 0.000625 + 0.050625 + 0.005625 + 0.030625 + 0.01 + 0 + 0.04 + 0.09 = 0.2275$$

$$\hat{\sigma}^2 = \frac{RSS}{n-k} = \frac{0.2275}{8-2} = 0.03792, \qquad \widehat{\mathrm{Var}}(\hat{\beta}_1) = \hat{\sigma}^2 \times 0.50 = 0.01896$$

$$\mathrm{se}(\hat{\beta}_1) = \sqrt{0.01896} = 0.1377, \qquad t = \frac{0.375}{0.1377} = 2.72$$

$$R^2 = 1 - \frac{RSS}{S_{yy}} = 1 - \frac{0.2275}{0.50875} = 0.553$$

The critical t-value at 5% with 6 degrees of freedom is $t_{0.025,6} = 2.447$, so the effect is significant in this toy sample. Converting from logs:

$$\%\Delta Y = (e^{0.375} - 1) \times 100 = (1.455 - 1) \times 100 = 45.5\%$$

4.3 OLS results for the full model ($n = 400$)

The matrix formulas of Section 3.3 give the coefficient vector (in the order constant, CREDIT, lnLAND, lnLAB, lnFERT, EDUC, EXT, FEM, AGE):

$$\widehat{\boldsymbol{\beta}}' = [5.2238 \quad 0.4567 \quad 0.0298 \quad 0.1441 \quad 0.1258 \quad 0.0173 \quad 0.1136 \quad -0.0786 \quad -0.0011]$$

The estimates with standard errors are below. HC1 t-statistics are shown beside the conventional ones.

Variable	Coefficient	Std. error	t	p-value	HC1 t
Constant	5.2238	0.3167	16.49	<0.001	15.16
CREDIT	0.4567	0.06	7.61	<0.001	7.7
lnLAND	0.0298	0.0543	0.55	0.583	0.56
lnLAB	0.1441	0.0623	2.31	0.021	2.28
lnFERT	0.1258	0.0423	2.97	0.003	2.94
EDUC	0.0173	0.0088	1.97	0.05	1.82
EXT	0.1136	0.0604	1.88	0.061	1.79
FEM	−0.0786	0.0596	−1.32	0.188	−1.36
AGE	−0.0011	0.0025	−0.43	0.667	−0.42

The sums of squares are $RSS = 129.442$ and $TSS = 162.547$, so:

$$\hat{\sigma}^2 = \frac{129.442}{400-9} = \frac{129.442}{391} = 0.33105, \qquad \hat{\sigma} = \sqrt{0.33105} = 0.5754$$

$$R^2 = 1 - \frac{129.442}{162.547} = 0.2037, \qquad \bar{R}^2 = 1 - \frac{129.442/391}{162.547/399} = 1 - \frac{0.33105}{0.40739} = 0.1874$$

$$F(8,391) = \frac{(162.547 - 129.442)/8}{129.442/391} = \frac{33.105/8}{0.33105} = \frac{4.138}{0.33105} = 12.50 \quad (p < 0.001)$$

For the credit coefficient, the t-statistic and the 95% confidence interval (critical value $t_{0.025,391} = 1.966$) are:

$$t = \frac{0.4567}{0.0600} = 7.61$$

$$0.4567 \pm 1.966(0.0600) = 0.4567 \pm 0.1180 = (0.339,\ 0.575)$$

In percentage terms:

$$\%\Delta Y = (e^{0.4567} - 1) \times 100 = (1.5789 - 1) \times 100 = 57.9\%$$

So OLS says credit users have about 57.9% higher yield. Labour and fertiliser have the expected positive elasticities (0.144 and 0.126). Schooling is positive and just significant at 5%, and extension is positive and significant at 10%. Land, sex of the head and age are not significant.

4.4 Diagnostic tests on the OLS model

The OLS model passes the standard checks, so inference from its standard errors is reliable; the problem it faces is endogeneity, not these.

Test	Statistic	5% critical value	p-value	Conclusion
VIF (mean / maximum)	1.03 / 1.07	10	n/a	No multicollinearity
Breusch-Pagan LM, $\chi^2(8)$	13.5	15.51	0.096	Homoskedasticity not rejected
Jarque-Bera, $\chi^2(2)$	0.912	5.99	0.634	Residuals consistent with normality
RESET, F(2, 389)	0.471	3.02	0.624	No functional-form misspecification

The calculations behind the table are as follows. For the credit regressor, $R_j^2 = 0.0631$, so $VIF = 1/(1 - 0.0631) = 1.067$. The mean of the eight VIFs is $8.219/8 = 1.027$.

$$LM = nR_{aux}^2 = 400 \times 0.03374 = 13.50 < \chi_{0.05}^2(8) = 15.51$$

With residual skewness $S = -0.117$ and kurtosis $K = 3.003$:

$$JB = \frac{400}{6}\left[(-0.117)^2 + \frac{(3.003 - 3)^2}{4}\right] = 66.67 \times 0.01368 = 0.912 < \chi_{0.05}^2(2) = 5.99$$

The RESET test adds $\hat{y}^2$ and $\hat{y}^3$, which reduces the residual sum of squares from 129.442 to 129.129:

$$F = \frac{(129.442 - 129.129)/2}{129.129/(400 - 9 - 2)} = \frac{0.1565}{0.3320} = 0.471 < F_{0.05}(2,389) = 3.02$$

4.5 Endogeneity and the 2SLS estimates

First stage. Both instruments are strongly related to credit access. Each extra kilometre to the nearest lender lowers the probability of credit by 2.95 percentage points, and group membership raises it by 27.67 percentage points:

$$\hat{\pi}_8 = -0.0295 \ (\text{se} = 0.0041), \quad t = \frac{-0.0295}{0.0041} = -7.20$$

$$\hat{\pi}_9 = 0.2767 \ (\text{se} = 0.0442), \quad t = \frac{0.2767}{0.0442} = 6.26$$

The joint F-statistic on the two instruments uses $RSS_R = 91.982$ (without DIST and GROUP) and $RSS_U = 74.499$ (with them):

$$F_1 = \frac{(91.982 - 74.499)/2}{74.499/(400 - 10)} = \frac{8.7415}{0.19102} = 45.76 > 10$$

The first-stage R^2 is 0.2412. H_0^3 is rejected, so the instruments are not weak.

Endogeneity test. Adding the first-stage residual to the OLS equation gives $\hat{\rho} = 0.4066$ with standard error 0.1517:

$$t_\rho = \frac{\hat{\rho}}{\text{se}(\hat{\rho})} = \frac{0.4066}{0.1517} = 2.68, \qquad p = 0.008$$

H_0^2 is rejected: credit is endogenous, and OLS is inconsistent.

2SLS estimates. The matrix formula of Section 3.4 gives the coefficient vector

$$\hat{\boldsymbol{\beta}}_{2SLS}' = [5.1332 \quad 0.1274 \quad 0.0445 \quad 0.1700 \quad 0.1394 \quad 0.0247 \quad 0.1263 \quad -0.1318 \quad -0.0002]$$

with the following standard errors and tests:

Variable	Coefficient	Std. error	t	p-value
Constant	5.1332	0.3306	15.53	<0.001
CREDIT	0.1274	0.1428	0.89	0.373

lnLAND	0.0445	0.0566	0.79	0.432
lnLAB	0.17	0.0655	2.6	0.01
lnFERT	0.1394	0.0442	3.15	0.002
EDUC	0.0247	0.0096	2.58	0.01
EXT	0.1263	0.0629	2.01	0.045
FEM	−0.1318	0.0653	−2.02	0.044
AGE	−0.0002	0.0027	−0.07	0.943

The 2SLS error variance uses $\hat{\mathbf{u}}'\hat{\mathbf{u}} = 139.417$:

$$\hat{\sigma}^2_{IV} = \frac{139.417}{400 - 9} = 0.35656$$

The credit coefficient falls from 0.4567 to 0.1274, a reduction of

$$1 - \frac{0.1274}{0.4567} = 1 - 0.279 = 0.721 \quad (72\%)$$

Its t-statistic, implied yield effect and 95% interval are:

$$t = \frac{0.1274}{0.1428} = 0.89 \ (p = 0.373), \qquad \%\Delta Y = (e^{0.1274} - 1) \times 100 = 13.6\%$$

$$0.1274 \pm 1.966(0.1428) = 0.1274 \pm 0.2808 = (-0.153,\ 0.408) \ \Rightarrow \ (e^{-0.153} - 1,\ e^{0.408} - 1)$$
$$= (-14.2\%,\ +50.4\%)$$

Since $t = 0.89 < 1.966$, H_0^1 is not rejected at 5% under 2SLS.

Instrument validity. The Sargan statistic is

$$S = nR^2_{\hat{u}} = 400 \times 0.0012035 = 0.481 < \chi^2_{0.05}(1) = 3.841 \quad (p = 0.488)$$

so H_0^4 is not rejected. As a cross-check, a Hausman-type contrast on the credit coefficient gives the following, which also exceeds 3.841 ($p \approx 0.011$) and agrees with the Durbin-Wu-Hausman test:

$$H = \frac{(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS})^2}{\widehat{\mathrm{Var}}(\hat{\beta}_{2SLS}) - \widehat{\mathrm{Var}}(\hat{\beta}_{OLS})} = \frac{(0.1274 - 0.4567)^2}{0.1428^2 - 0.0600^2} = \frac{(-0.3293)^2}{0.01679} = \frac{0.1084}{0.01679} = 6.46$$

4.6 Comparing OLS and 2SLS, and discussion

Estimator	Credit coefficient	Std. error	95% CI	Implied yield effect
OLS	0.4567	0.06	(0.339, 0.575)	57.90%
2SLS	0.1274	0.1428	(−0.153, 0.408)	13.60%
True value in simulation	0.18	n/a	n/a	19.70%

Measured against the true value $\beta_1 = 0.18$, the estimation errors are:

$$\hat{\beta}_{OLS} - \beta_1 = 0.4567 - 0.18 = 0.2767, \qquad \hat{\beta}_{2SLS} - \beta_1 = 0.1274 - 0.18 = -0.0526$$

The true yield effect is $(e^{0.18} - 1) \times 100 = 19.7\%$. The OLS estimate overshoots, and the true value lies outside its confidence interval. The 2SLS estimate is close, and the true value lies inside its interval. This is the pattern predicted by positive selection: unobserved ability raises both the chance of obtaining credit and the yield, so OLS attributes ability's effect to credit.

The price of removing the bias is precision:

$$\frac{\text{se}(\hat{\beta}_{2SLS})}{\text{se}(\hat{\beta}_{OLS})} = \frac{0.1428}{0.0600} = 2.38$$

The 2SLS standard error is 2.38 times the OLS standard error, because only the variation in credit that is driven by the instruments is used. The 2SLS result should therefore not be read as proof that credit has no effect: its interval ranges from a 14% loss to a 50% gain in yield.

Hypothesis	Test	Statistic	p-value	Decision at 5%
H1: no credit effect	t-test, OLS	7.61	<0.001	Reject
H1: no credit effect	t-test, 2SLS	0.89	0.373	Fail to reject
H2: credit exogenous	Durbin-Wu-Hausman t	2.68	0.008	Reject
H3: instruments irrelevant	First-stage F	45.76	<0.001	Reject
H4: instruments valid	Sargan $\chi^2(1)$	0.481	0.488	Fail to reject

Two interpretive limits apply. First, the 2SLS coefficient is a local average treatment effect: it describes farmers whose credit status changes with distance to a lender or group membership, not every farmer. Second, the exclusion restriction cannot be tested directly; the Sargan test only checks consistency between the two instruments.

5.0 CONCLUSION AND POLICY RECOMMENDATIONS

5.1 Conclusion

Credit looks far more productive in a naive regression than it is once selection is addressed. In the illustrative sample, OLS puts the credit effect at 0.4567 log points (+57.9% yield), while 2SLS puts it at 0.1274 (+13.6%) with a 95% interval that spans zero. The first-stage F-statistic of 45.76 and the Sargan p-value of 0.488 support the instruments, and the Durbin-Wu-Hausman test ($p = 0.008$) rejects the exogeneity of credit. Against the known true value of 0.18, the OLS estimate is biased upward and the 2SLS estimate is close. The main lesson for credit evaluation is that the size of a credit effect cannot be read from a comparison of borrowers and non-borrowers.

5.2 Policy recommendations

1. **Evaluate credit programmes with credible identification.** Counties and lenders should use randomised or phased roll-outs, or valid instruments, instead of borrower versus non-borrower comparisons, which overstate impact.
2. **Lower the cost of reaching a lender.** Distance has a strong negative effect on credit access, so agent banking, mobile-based lending and SACCO outreach in remote wards are likely to widen access at low cost.
3. **Use group channels.** Group membership raises the probability of credit by about 28 percentage points, which supports group lending and loan guarantees through farmer groups and SACCOs.
4. **Bundle credit with complementary inputs and advice.** Fertiliser, labour and extension contact all carry positive coefficients, so credit that finances these inputs together with extension services is more likely to raise yield than cash alone.

5. **Pair credit with risk management.** Evidence from Ghana (Karlan et al., 2014) suggests uninsured risk limits what farmers do with credit; index insurance or flexible repayment can complement lending.
6. **Build better data.** KNBS and research institutions should maintain household panels that record credit, distance to lenders and group membership, so that causal effects can be estimated on Kenyan data.

5.3 Limitations and areas for further research

- The data are simulated, so the numerical findings illustrate the method and are not evidence about Kenya. The next step is to re-estimate the model on survey data.
- The first stage is a linear probability model; a probit first stage or a bivariate-probit specification could be compared.
- The exclusion restriction for GROUP may fail if groups also pass on extension advice or inputs. Using DIST alone would be just-identified and would remove the Sargan test, so the choice of instruments should be defended with institutional detail.
- The Cobb-Douglas form is restrictive. A translog specification, or stochastic frontier analysis of technical efficiency, would allow credit to affect efficiency directly.
- A single cross-section cannot capture dynamic effects. Panel data with farm fixed effects would remove time-invariant ability.

REFERENCES

1. Aguinis, H., Pierce, C. A., Bosco, F. A., & Muslin, I. S. (2009). First decade of *Organizational Research Methods*: Trends in design, measurement, and data-analysis topics. *Organizational Research Methods, 12*, 69–112. https://doi.org/10.1177/1094428108322641
2. Bao Hong, T. (2008). *Cobb-Douglas production function*. http://docentes.fe.unl.pt/~jamador/Macro/cobb-douglas.pdf
3. Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica, 47*(5), 1287–1294.
4. Carter, M. R. (1989). The impact of credit on peasant productivity and differentiation in Nicaragua. *Journal of Development Economics, 31*(1), 13–36.
5. Conning, J., & Udry, C. (2007). Rural financial markets in developing countries. In R. Evenson & P. Pingali (Eds.), *Handbook of agricultural economics* (Vol. 3). Elsevier.
6. Creswell, J. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education.
7. Feder, G., Just, R. E., & Zilberman, D. (1985). Adoption of agricultural innovations in developing countries: A survey. *Economic Development and Cultural Change, 33*(2), 255–298.
8. Feder, G., Lau, L. J., Lin, J. Y., & Luo, X. (1990). The relationship between credit and productivity in Chinese agriculture: A microeconomic model of disequilibrium. *American Journal of Agricultural Economics, 72*(5), 1151–1157.
9. Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
10. Government of Kenya. (2019). *Agricultural Sector Transformation and Growth Strategy 2019–2029*. Ministry of Agriculture, Livestock, Fisheries and Cooperatives.
11. Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics, 2*(2), 111–120.
12. Greene, W. H. (2018). *Econometric analysis* (8th ed.). Pearson.
13. Guala, F., & Mittone, L. (2005). Experiments in economics: External validity and the robustness of phenomena. *Journal of Economic Methodology, 12*(4), 495–515.
14. Gujarati, D. N. (2004). *Basic econometrics*. McGraw-Hill.
15. Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica, 46*(6), 1251–1271.

16. Hoff, K., & Stiglitz, J. E. (1990). Imperfect information and rural credit markets: Puzzles and policy perspectives. *World Bank Economic Review, 4*(3), 235–250.
17. Ingado Aluvisia, J., & Kuso, Y. (2025). Credit access and growth of women-owned microenterprises at Kibuye Market, Kisumu County, Kenya. *EPRA International Journal of Economics, Business and Management Studies*.
18. Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review, 55*(2), 163–172.
19. Johansen, A. (2011). *Determinants of capital structure during credit bubble and credit crunch: An empirical investigation of Danish SMEs* [Unpublished bachelor's thesis]. University of Aarhus, Denmark.
20. Karlan, D., Osei, R., Osei-Akoto, I., & Udry, C. (2014). Agricultural decisions after relaxing credit and risk constraints. *Quarterly Journal of Economics, 129*(2), 597–652.
21. Kuso, Y. (2020). Research of human capital impact on economic growth in Kenya. *JSES International*.
22. Kuso, Y., & Gachunga Muhia, J. (2019a). Agricultural exports and food insecurity in Sub-Saharan Africa: A qualitative configurational analysis. *International Journal of Management and Economics*.
23. Kuso, Y., & Gachunga Muhia, J. (2019b). Agricultural exports and food insecurity in Sub-Saharan Africa: A qualitative configurational analysis. *International Journal of Scientific Research and Management Studies*.
24. Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica, 55*(3), 703–708. https://doi.org/10.2307/1913610
25. Ochieng, G. J., & Kuso, Y. (2026). Role of government subsidies on agricultural productivity in Migori County, Kenya. *International Journal on Research and Development – A Management Review*.
26. Orina, S. A., & Kuso, Y. (2025). Inflation dynamics and economic growth in developing economies: Evidence from Kenya, Nigeria, Ethiopia, and Ghana (2008–2024). *IJESS International Journal of Education and Social Science*.
27. Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics, 16*(3), 289–326.
28. Petrick, M. (2005). Empirical measurement of credit rationing in agriculture: A methodological survey. *Agricultural Economics, 33*(2), 191–203.
29. Phillips, P. C. B., & Perron, P. (1988). Testing for unit roots in time series regression. *Biometrika, 75*(2), 335–346.
30. Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society, Series B, 31*(2), 350–371.
31. Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica, 26*(3), 393–415.
32. Scandura, T. A., & Williams, E. A. (2000). Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal, 43*(6), 1248–1264.
33. Ssengonzi, J., Kuso, Y., & Olyanga, A. M. (2025). Contribution of social capital and microcredit accessibility on economic welfare of small-scale farmers in Mityana District, Uganda. *East African Journal of Business and Economics*.
34. Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica, 65*(3), 557–586.
35. Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. *American Economic Review, 71*(3), 393–410.
36. White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test of heteroskedasticity. *Econometrica, 48*(4), 817–838.
37. Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.
38. Wu, D.-M. (1973). Alternative tests of independence between stochastic regressors and disturbances. *Econometrica, 41*(4), 733–750.