Technical Efficiency of Kenyan Smallholder Dairy Farmers in Different Agro-ecological Regions

Vincent Ngeno (vcngen0@gmail.com)
Moi University

Chepchumba Chumo
Moi University

Research Article

Keywords: smallholder dairy sector, stochastic meta-frontier, panel data, Kenya

Posted Date: September 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-848912/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Technical Efficiency of Kenyan Smallholder Dairy Farmers in Different Agro-ecological Regions

Vincent Ngeno
Department of Agricultural Economics & Resource Management
Moi University
P.O Box 3900 – 30100
Eldoret
Email: vcngeno@gmail.com
Telephone: +254 721 578 506
ORCID: 0000-0003-1833-6840

Chepchumba Chumo
Department of Agricultural Economics & Resource Management
Moi University
P.O Box 3900 – 30100
Eldoret
Email: chumbiin@gmail.com
Telephone: +254 777 770 770

Corresponding author: Email: vcngeno@gmail.com / Telephone: +254 721 578 506

Technical Efficiency of Kenyan Smallholder Dairy Farmers in Different Agro-ecological Regions

Abstract

We employ stochastic meta-frontier and region-specific frontiers based on the “true” random effect framework to examine technical efficiencies, technology gaps and meta-frontier technical efficiency of Kenyan smallholder dairy farmers in different agro-ecological zones. The empirical analysis is based on comprehensive three-wave household level panel data from across three agro-ecological zones in Kenya. Results show variations in efficiency measures and that smallholder milk production is characterized by increasing returns scale across all agro-ecological zones. The results indicate that the milk output of smallholder dairy farmers from all the agro-ecological zones lags behind their potential with the technology available and the prevailing environment conditions in their agro-ecological zones. We also, find that there exists a significant technology gap in dairy production across the agro-ecological zones in Kenya. These findings generate important policy implication for achieving increased technical efficiency and reduce the technology gap in smallholder dairy production.

Keywords: smallholder dairy sector, stochastic meta-frontier, panel data, Kenya

JEL Classification: C21, C23, Q12, O13

1. Introduction

The dairy sector in Kenya is one of the largest agricultural segments of the country contributing about 40% of the livestock Gross Domestic Product (GDP), 14% of the agricultural GDP, and 4% of the overall GDP (KDB, 2021). With a dairy cattle population of 4.6 million and annual milk production of 5 billion liters in 2019, Kenya is the largest milk producer in Africa. Milk consumption rates in Kenya are among the highest in sub-Saharan Africa: between 50 and 150L per year (Faostat, 2021). The country’s annual per capita consumption of milk has recently been estimated at 110kg, this falls short of the widely recommended 220kg annual per capita consumption (Muunda et al., 2021). About 80% of the milk in Kenya is produced by and marketed by smallholder dairy farmers (SDFs) (Bateki, van Dijk, Wilkses, Dickhoefer, & White, 2020). McDermott, Staal, Freeman, Herrero, and Van de Steeg (2010) described the Kenyan SDFs as small farms often comprising less than 5 ha land, keeping one to five dairy cows that are often improved breed (Holstein, Friesian, or Ayrshire mixed with local breeds). The average daily milk production per cow in most smallholder dairy farms is estimated to be about 6-7 liters/cow/day which is approximately 70% lower than that of cows in developed countries (Makau et al., 2020). Such a low production level, compared to that of global players, represents a huge economic loss.

The dairy sector is a source of livelihood and provides food, income and employment for many millions of people in Kenya. Growing at an annual rate of about 5 to 7%, the sector is a source of livelihood to roughly 1.8 million SDFs who account for over 80% of the country’s milk producers (KDB, 2018). Although the dairy sector is expanding in Kenya, the rate of growth does not match the increased demand for dairy products being experienced in the country and beyond. As the human population increases, incomes rise and urbanization continues, this deficit in domestic milk supply will become a food insecurity challenge. Thus, increasing milk yields of SDFs is a priority for enhancing the well-being of consumers, producers and their families. Modeling studies have projected an 85% increase in milk demand in Kenya between 2010 and 2050, while the government’s dairy policy aims to increase milk production by 130% by 2030 (Wilkes, Wassie, Odhong, Fraval, & van Dijk, 2020). To meet the milk demand, milk production in Kenya ought to increase.
The Kenyan government encourages SDFs to produce more milk by increasing farm productivity and efficiency. In order to have a better understanding of the scope for increasing the output of milk in the country, it is imperative to have some knowledge of the productive capacity of SDFs. Gelan and Muriithi (2012) and Maina, Mburu, Gitau, and VanLeeuwen (2020) have examined technical efficiency of smallholder milk production in Kenya. A common feature of these studies is that they are based on a hypothesis that all the SDFs share a common technology. However, Kenya has heterogeneous agro-ecological conditions which would affect milk productive efficiency. Hence, the hypothesis of homogenous technology may not apply to the SDFs of Kenya. It is more plausible to assume that different types of agro-ecological regions in Kenya may employ different types of technology. However, to our knowledge, almost no studies have taken this into account when evaluating the technical efficiency of smallholder milk production in the agro-ecological regions of Kenya. Ignoring the variation of technologies in Kenya’s agro-ecological regions may lead to biased estimates of SDFs efficiency scores, thereby misleading policy implications.

Smallholder dairy production areas span a wide agro-ecological gradient, including semi-arid, sub-humid and temperate highlands, which differ in terms of altitude, soil types, landscapes, and climatic conditions. The geographical and climatic variations lead to different levels of resource endowment and production potentials which causes further variations in farming systems and socio-economic conditions. The aforementioned factors have led to wide heterogeneity in the available production technology set for the farming households in different agro-ecological regions. Therefore, these differences in agro-ecological regions potentially inhibit SDFs in some regions from choosing the best technology from a set of potential technologies resulting in the so-called technology gap ratios (TGRs). In this study, we deal with heterogeneous technologies through integrating their measurement of SDFs efficiency with the parametric meta-frontier approach first proposed by Battese and Rao (2002). This method enables us to estimate technology gaps and comparable efficiencies for milk production under distinct technologies relative to the potential technology available to the SDFs.

The objective of this study is to provide new evidence on the technology gap and SDFs efficiency performance in Kenya using a parametric meta-frontier approach. The remainder of this study is organized as follows. In Section 2, we introduce the parametric meta-frontier model to measure the SDFs efficiency and estimate the technology gap. In Section 3, we present the empirical analysis of the case of SDFs in Kenya. In Section 4, we conclude the study.

2. Methodology

2.1 Analytical strategy

To consider the heterogeneities of agro-ecological regions, we need to employ the meta-frontier analysis method. The meta-frontier analysis is deeply rooted in production theory. It differentiates the technology heterogeneities by estimating region-specific production frontier. It then estimates the meta-frontier by searching for the envelope of the region-specific frontiers. Suppose we have \( k \) agro-ecological regions in Kenya, we can then estimate group stochastic frontiers for each region as follows:

\[
y_{it}^k = f^k(x_{it}^k, \beta^k)e^{(v_{it}^k-u_{it}^k)} i = 1, 2, ... , N(k)
\]

where \( y_{it}^k \) denotes the output level for SDF \( i \) in the \( k \)th region in the \( ith \) time period, \( x_{it}^k \) is the input vector, \( v_{it}^k \) represents the error term and is assumed to be iid as \( v_{it}^k \sim N(0, \sigma_{vk}^2) \). \( u_{it}^k \) is a one-sided error representing technical inefficiency and is distributed as \( u_{it}^k \sim N^+(0, \sigma_{vk}^2(z_{it}^k)) \), where \( z_{it}^k \) denotes
inefficiency or production environment determinants, and \( \beta^k \) is a vector of unknown parameters for the \( k \)th region. The technical efficiency (TE) of the \( i \)th SDF relative to the region \( k \) frontier can be computed, following Greene (2005b), as:

\[
TE^k_{it} = \frac{y^k_{it}}{f^k (x^k_{it}, \beta^k) e^{(u^k_{it})}} = \frac{f^k (x^k_{it}, \beta^k) e^{(-u^k_{it})}}{f^k (x^k_{it}, \beta^k)} = e^{-u^k_{it}}
\]  

(2)

where \( TE^k_{it} \) is a measure of the performance of the individual SDF (i) relative to the regional group frontier.

To estimate the stochastic meta-frontier function that envelopes all the frontiers of the \( k \) agro-ecological regions, we use the approach by Battese, Rao, and O’donnell (2004) defined as

\[
Y^*_{it} = f(x_{it}, \beta^*) \equiv e^{x_{it}^\beta^*}, i = 1, 2, \ldots, N, N = \sum_{i=1}^{N} N_j, t = 1, \ldots, T
\]  

(3)

where \( \beta^* \) is the vector of unknown meta-frontier parameters. The meta-frontier should be larger than or equal to the group-specific frontier, that is, \( x_{it}^\beta^* \geq x_{it}^\beta_j \). Equation (3) indicates that the meta-frontier analysis production function contains all the group frontiers over the entire period. For simplicity, the function \( f \) in Equation (1) is assumed to be \( e^{x_{it}^\beta(j)} \); then \( e^{x_{it}^\beta(j)} \); Equation (1) can be written as follows:

\[
Y_{it} = e^{-u_{it}(j)} \times e^{x_{it}^\beta(j)} = e^{x_{it}^\beta^* + V_{it}(j)}
\]  

(4)

By dividing Equation (5) by \( e^{x_{it}^\beta^* + V_{it}(j)} \), we obtain

\[
\frac{Y_{it}}{e^{x_{it}^\beta^* + V_{it}(j)}} = e^{-u_{it}(j)} \times \frac{e^{x_{it}^\beta(j)}}{e^{x_{it}^\beta^*}}
\]  

(5)

By definition, \( e^{-u_{it}(j)} \) is the TE of group \( j \), and the second part of Equation (5) is the ratio of the \( i \)th frontier function to the meta-frontier production function, which is called the technology gap ratio (TGR).

The estimated TGR must be less than or equal to unity. The great virtue of the TGR estimator is that establishments with heterogeneous production functions can be compared in terms of relative efficiency. The TE of the meta-frontier production function, \( TE^* \), is expressed as a product of TE and TGR:

\[
TE^*_{it} = \frac{Y_{it}}{e^{x_{it}^\beta^* + V_{it}(j)}} = TE_{it} \times TGR_{it}
\]  

(6)

The TE component captures the performance of the SDFs compared with the best-performed SDF in the sample, while TGR components capture the differences in frontier or best-practice technology in the group. Their product (\( TE^* \)) defines the technical efficiency of SDFs by meta-frontier.

The TE of individual SDF is estimated assuming a flexible translog functional form Equation (2) and (3), which accounts for nonlinearity, substitution, and complementarity, as well as non-constant returns to scale. The model is estimated using the ‘true’ random effect (TRE) frontier model (Greene, 2005b), and it extends the conventional stochastic frontier model by disentangling the farm effect (unobserved heterogeneity) from TE. The trend variable, \( t \), is introduced to capture the effect of technological change. The estimated flexible translog function for the region \( k \) frontier is specified as;
\[ lny_{it}^k = \beta_0^k + \sum_{j=1}^{4} \beta_j^k \ln x_{jit} \]
\[ + \frac{1}{2} \sum_{j=1}^{4} \beta_{jj}^k (\ln x_{jit})^2 + \sum_{j=1}^{4} \sum_{l=1}^{4} \beta_{jl}^k \ln x_{jit} \ln x_{ilt} + \beta_t^k t + \frac{1}{2} \beta_{tt}^k \]
\[ + \sum_{j=1}^{4} \beta_{j1}^k \ln x_{jit} t + \theta_i^k + v_i^k - u_i^k \]

where \( y_{it} \) is a vector of dairy outputs, \( x_{jit} \) is a vector of inputs \((j = 1, \ldots, J)\) by farms \((i = 1, \ldots, N)\) over time \((t = 1, \ldots, T)\), and all the Greek letters are parameters to be estimated. The white-noise error term \( v_{it} \) is added to allow for random measurement error. The term \( v_{it} \) is symmetrical and is assumed to satisfy the classical assumptions, that is, \( v_{it}^{id} \sim N(0, \sigma_v^2), v_{it} \perp u_{it} \). The term \( u_{it}^k \) is specified as \( u_{it}^k \sim N^+ \left(0, \sigma_{u_{it}}^2 \right)\), and \( \theta_i^k \) is a farm-specific component for capturing time-invariant unobserved heterogeneity, which is assumed to have an iid normal distribution.

### 2.2 Data and definition of variables

We use a household-level panel dataset collected in 2014, 2017 and 2020, comprising of 1344 households across three agro-ecological zones in Kenya. The agro-ecological zones include lower highland (LH), upper midland (UM) and lower midland (LM). LH has a mean temperature of 15-18°C, an altitude range of 1800-2400 m above sea level (asl). Kericho County covered LH1, which is moderate and humid, with an annual average rainfall of >80% of the potential evaporation. UM has a mean temperature of 18-21°C and an altitude range of 1300-1900 m asl. Bomet County covered UM1 and UM2. UM1 is temperate and humid, with an annual average rainfall of >85% of the potential evaporation, UM2 is temperate and subhumid, with an annual average rainfall of 65-68% of the potential evaporation. LM has a mean temperature of 21-24°C and an altitude range of 800 to 1500 m asl. Narok County covered LM1, which is temperate and semiarid, with an annual average rainfall of 25-40% of the potential evaporation.

The sampling procedure involved both purposive and random sampling techniques. The three agro-ecological regions and three counties were selected purposively, considering their potential smallholder dairy production. The main purpose of the survey was to generate information on current dairy farming practices, technology adoption, access to institutional services, and prevailing production constraints in smallholder dairy production systems in the country. Following the identification of agro-ecological regions and counties, a proportional to size sampling procedure was used to select 3-6 wards (smallest administrative division in Kenya) from each county in which 84-92 SDFs were identified for interviews.

Finally, 1512 randomly selected respondents were interviewed during the first-round survey and 1444 and 1354 were re-interviewed in the second and third rounds respectively using the same questionnaire. In the final analyses, the balanced data of 1344 respondents, with an attrition level of less than 5%, were used. A structured questionnaire was prepared and pre-tested on selected respondents for further modification to ensure the validity of all questions so that the required SDFs information could be captured properly. The questionnaires were administered by experienced enumerators who underwent specific training on the questionnaires before execution of the survey. Finally, the survey was executed under the close supervision of researchers from Moi University.
thorough data cleaning exercise was carried out before using the Stata 16 software program for the data analysis.

The variables considered in both $x_{ijt}$ and $z_{it}$ in Equation (2) and (3) are selected based on a review of related efficiency literature (Alem, Lien, Hardaker, & Guttormsen, 2019; Girma, 2019; Ma, Bicknell, & Renwick, 2019; Mareth, Scavarda, Thomé, Oliveira, & Alves, 2019; Mareth, Thomé, Scavarda, & Oliveira, 2017; Mehmood, Rong, Bashir, & Arshad, 2018; Sabasi, Shumway, & Astill, 2019; Skevas, Emvalomatis, & Brümmer, 2018). The input variables $x_{ijt}$s include: $x_1$ is farmland, defined as productive land (both owned and rented) in hectares, $x_2$ is labour, measured as total labour hours used on the farm, including hired labour, owners' labour, and family labour, $x_3$ is the number of milk cows measured as livestock units, $x_4$ is capital, a service flow from land and buildings estimated as 5 percent of market value plus accrual machinery hire expenses, accrual machinery repair expenses, and machinery depreciation, $x_5$ is accrual purchased feed measured in Kenya Shillings (KES), and $x_6$ is other expenses which includes breeding expenses, veterinary and medicines, and other livestock expenses (KES).

The $z_{it}$-variables in this study consist of the following: $z_1$ is number of years of schooling achieved by the household head, $z_2$ is age of household head in years, $z_3$ is the age squared to control for the non-linear life cycle effects, $z_4$ is a dummy variable that takes a value of one if the household head has access to credit, otherwise zero, $z_5$ is a dummy variable that takes a value of one if household participated in off-farm work, 0 otherwise, $z_6$ is a dummy variable that takes a value of one if a farmer participated in a dairy farmer group, 0 otherwise and $z_7$ is a dummy variable that equals one if a farmer received extension service and zero otherwise.

Table 1

| Variable | Label       | LH Mean | LH Standard deviation | UM Mean | UM Standard deviation | LM Mean | LM Standard deviation |
|----------|-------------|---------|-----------------------|---------|-----------------------|---------|-----------------------|
| Production function variables | | | | | | | |
| $y$      | Milk yield  | 1295    | 173.53                | 1166    | 182.77                | 1084    | 169.35                |
| $x_1$    | Farmland    | 2.9     | 2.1                   | 3.5     | 2.9                   | 3.7     | 2.8                   |
| $x_2$    | Labour      | 2798    | 1148                  | 2876    | 1259                  | 2679    | 1138                  |
| $x_3$    | Cows        | 2.9     | 2.7                   | 3.6     | 2.6                   | 3.1     | 2.9                   |
| $x_4$    | Capital     | 85532   | 91828                 | 83242   | 102716                | 80111   | 80504                 |
| $x_5$    | Feeds       | 168557  | 206525                | 146376  | 217413                | 113043  | 199101                |
| Inefficiency determinants variables | | | | | | | |
| $z_1$    | Education   | 8.31    | 3.83                  | 7.29    | 4.82                  | 6.79    | 4.37                  |
| $z_2$    | Age         | 48.87   | 18.25                 | 48.89   | 19.41                 | 49.54   | 17.36                 |
| $z_3$    | Credit      | 0.25    | 0.17                  | 0.34    | 0.25                  | 0.23    | 0.19                  |
| $z_4$    | Off-farm    | 0.39    | 0.19                  | 0.58    | 0.37                  | 0.48    | 0.28                  |
| $z_5$    | Group       | 0.46    | 0.49                  | 0.47    | 0.45                  | 0.51    | 0.35                  |
| $z_6$    | Extension   | 0.54    | 0.38                  | 0.32    | 0.39                  | 0.29    | 0.40                  |
| Number of observations | 452        | 445                 | 447                  | | | | |
Descriptive statistics of the variables used in the study are reported in Table 1. A quick view of Table 1 shows that considerable differences exist among the three agro-ecological zones in terms of mean input and milk output. The highest average milk output was recorded in the LH zone and followed by UM zone while the LM zone recorded the lowest milk output in the period considered. In terms of farm inputs, the LH zone has, on average, the highest use of capital and the highest use of dairy feeds. Similarly, the UM zone has the highest number of dairy cows and the highest level of labour use. SDFs in the LM zone have on average the highest level of land under dairy. Under SDF-specific variables, it can be observed that SDFs in the LH zone have, on average, the best level of education levels and highest access to extension services. In terms of access to credit and off-farm income, the UM zone has recorded the highest percentage. The LM zone has significantly older SDFs and the highest percentage of associated with farmers groups.

3. Results and discussion

Before estimating the empirical model, the variables measured in monetary values were deflated, input variables were normalized. Given that we had no agro-ecological zone-level price indices in the current data set, we compute the real values using the 2020 agro-ecological zonal consumer prices indices published by Kenya Bureau of Statistics to deflate the series. We normalized the variables by dividing values by their respective geometrics means. Four hypothesis tests were carried out to verify our estimations' validity. The results of the four hypotheses were tested using the generalized likelihood-ratio test statistic. The results of the four null hypotheses tested are reported in Table 2.

**Table 2**

Results of testing hypothesis

| Hypothesis                                                                 | LR Statistics | LR Critical Value | Decision         |
|---------------------------------------------------------------------------|---------------|-------------------|------------------|
| 1: There are no inefficiency effects                                        |               |                   |                  |
| \( H_0: \delta_{jks} = 0, \gamma = 0 \)                                  |               |                   |                  |
| LH                                                                        | 4.86          | 2.71              | Reject \( H_0 \) at 5% |
| UM                                                                        | 7.17          | 5.41              | Reject \( H_0 \) at 1% |
| LM                                                                        | 8.83          | 5.41              | Reject \( H_0 \) at 1% |
| Pooled                                                                    | 19.60         | 5.41              | Reject \( H_0 \) at 1% |
| 2: The coefficients of the square values and the interaction terms in translog have zero values |               |                   |                  |
| \( H_0: \beta_{ij}^k = 0 \)                                              |               |                   |                  |
| LH                                                                        | 33.28         | 30.58             | Reject \( H_0 \) at 1% |
| UM                                                                        | 28.13         | 25.00             | Reject \( H_0 \) at 5% |
| LM                                                                        | 29.02         | 25.00             | Reject \( H_0 \) at 5% |
| Pooled                                                                    | 36.66         | 30.58             | Reject \( H_0 \) at 1% |
| 3: Exogenous factors are not responsible for the inefficiency term          |               |                   |                  |
| \( H_0: \delta_{jks} = 0 \)                                              |               |                   |                  |
| LH                                                                        | 24.11         | 17.76             | Reject \( H_0 \) at 1% |
| UM                                                                        | 15.75         | 13.40             | Reject \( H_0 \) at 5% |
| LM                                                                        | 17.57         | 13.40             | Reject \( H_0 \) at 5% |
| Pooled                                                                    | 82.17         | 17.76             | Reject \( H_0 \) at 1% |
| 3: Stochastic frontiers of different agro-ecological zones are identical.  |               |                   |                  |
| \( H_0: f(X_{ki}; \beta_k) = f(X_{ji}; \beta_j) \)                       |               |                   |                  |
| Same agro-ecological zone frontier                                         | 229.65        | 128.80            | Reject \( H_0 \) at 1% |
The first null hypothesis of no technical inefficiency in the agro-ecological zonal production frontier and the pooled data was rejected for each agro-ecological zone and the pooled data. Hence, the traditional average response model is inappropriate for the data set, given the assumption of the stochastic frontier model. The second null hypothesis that the second-order coefficients of the translog model have zero values was also rejected for each agro-ecological zone and the pooled data. Thus, the translog model gives more accurate and consistent results than Cobb-Douglas model, given the data set. The third null hypothesis that technical inefficiency effects in the stochastic frontier model are not explained by any of the covariates in the inefficiency model was also rejected for all agro-ecological zones and pooled data. Thus, SDFs’ inefficiency is influenced by socio-economic, environmental and farm-specific characteristics.

The final null hypothesis that all the agro-ecological zones share the same technology was considered. If the three agro-ecological zones were to share the same production frontier, there would be no good reason for estimating the efficiency levels of groups relative to a meta-frontier production function. According to Battese et al. (2004), there would be no good reason for estimation of technical efficiency of SDFs relative to meta-frontier if all the data were generated from a single production frontier. Fourth null hypothesis is also rejected. This result suggests that there is evidence to support the fact that SDFs operate under heterogeneous technologies and therefore, the three agro-ecological zones stochastic frontiers for SDFs in Kenya are different, a result which justifies the use of the meta-frontier model for a comparison of technical efficiency.

**Stochastic regional frontiers and meta-frontier estimates**

The technical efficiency estimates were obtained from the stochastic frontier model and estimated separately for the three agro-ecological zones and the pooled data following the procedure for the “true” random effect model using STATA version 16. The meta-frontier estimated using SHAZAM version 11 following O’Donnell, Rao, and Battese (2008). The estimated parameters of the flexible translog function specified in Equation (7) for estimation for the agro-ecological zones and the pooled are reported in Table 3. Table 3 also shows the results of the linear programming estimates for the meta-frontier.

The sigma-squared value of 0.55 in LH, 0.22 in UM, 0.20 in LM and 0.26 in all the agro-ecological zones is evidence that the model explains considerable variation in output. The gamma value which estimates the deviation of the observed output from the frontier output are 0.87, 0.66, 0.78 and 0.77 for LH, UM, LM and in all the agro-ecological zones, respectively. This suggests that a larger proportion of the deviation in total output for LH, UM, LM and in all the agro-ecological zones could be attributed to inefficiency in the use of inputs and other farm practices. However, the stochastic noise such as unfavorable weather conditions and data measurement errors contribute a relatively smaller proportion of the deviation in output from the frontier.

| Parameter | LH         | UM         | LM         | Pooled     | Meta-frontier |
|-----------|------------|------------|------------|------------|---------------|
| \( \beta_0 \) | 2.387***   | 2.638***   | 3.885***   | 2.857***   | 2.583***      |
|           | (0.749)    | (0.673)    | (1.238)    | (0.788)    | (0.576)       |
\[
\begin{array}{cccccc}
\beta_1 & 0.212^{**} & -0.037 & -0.456 & 0.384^{***} & 0.167 \\
& (0.085) & (0.049) & (0.265) & (0.131) & (0.135) \\
\beta_2 & -0.225 & 0.172 & 0.296^{***} & 0.562^{**} & 0.285^{**} \\
& (0.322) & (0.467) & (0.063) & (0.242) & (0.142) \\
\beta_3 & 0.428^{**} & 0.336^{*} & 0.469^{***} & 0.311^{***} & 0.249^{*} \\
& (0.191) & (0.176) & (0.135) & (0.098) & (0.131) \\
\beta_4 & 0.121^{**} & 0.036 & 0.052 & 0.514^{***} & 0.155 \\
& (0.055) & (0.027) & (0.039) & (0.148) & (0.129) \\
\beta_5 & 0.183 & 0.419^{**} & 0.169 & 0.687^{***} & -0.352 \\
& (0.211) & (0.189) & (0.369) & (0.194) & (0.252) \\
\beta_{1^2} & -0.288^{***} & -0.032 & -0.016^{*} & -0.044 & 0.079 \\
& (0.083) & (0.037) & (0.009) & (0.388) & (0.077) \\
\beta_{2^2} & -0.072 & 0.121^{***} & -0.043 & -0.102 & 0.477^{***} \\
& (0.047) & (0.034) & (0.028) & (0.063) & (0.167) \\
\beta_{3^2} & -0.191 & 0.289 & -0.151 & 0.091^{***} & -0.085 \\
& (0.217) & (0.181) & (0.123) & (0.021) & (0.069) \\
\beta_{4^2} & 0.063^{**} & 0.029 & 0.152^{**} & -0.082 & 0.075 \\
& (0.027) & (0.027) & (0.068) & (0.052) & (0.062) \\
\beta_{5^2} & 0.156 & 0.117^{**} & 0.110 & 0.194^{**} & 0.079^{**} \\
& (0.121) & (0.047) & (0.081) & (0.061) & (0.031) \\
\beta_{12} & 0.037 & 0.274^{***} & 0.065 & 0.234^{***} & 0.049 \\
& (0.056) & (0.092) & (0.060) & (0.023) & (0.039) \\
\beta_{13} & -0.035 & -0.054^{**} & -0.013 & -0.041 & 0.253^{***} \\
& (0.032) & (0.022) & (0.011) & (0.051) & (0.082) \\
\beta_{14} & 0.043^{***} & -0.026 & -0.014 & 0.034^{***} & -0.063 \\
& (0.013) & (0.023) & (0.010) & (0.013) & (0.047) \\
\beta_{15} & -0.031 & -0.302 & -0.603^{***} & -0.328^{***} & -0.315^{***} \\
& (0.143) & (0.274) & (0.121) & (0.101) & (0.074) \\
\beta_{23} & -0.127 & -0.037^{**} & -0.043 & -0.003 & -0.007 \\
& (0.113) & (0.016) & (0.027) & (0.031) & (0.025) \\
\beta_{24} & 0.051^{***} & -0.031 & -0.032 & -0.033^{***} & -0.063 \\
& (0.006) & (0.025) & (0.042) & (0.001) & (0.046) \\
\beta_{25} & 0.022 & -0.006 & 0.046^{**} & 0.043 & 0.031 \\
& (0.024) & (0.049) & (0.018) & (0.035) & (0.041) \\
\beta_{34} & -0.040 & 0.143^{***} & -0.024 & 0.079^{**} & 0.321^{***} \\
& (0.031) & (0.014) & (0.015) & (0.039) & (0.056) \\
\beta_{35} & 0.068^{***} & -0.421 & 0.187^{***} & 0.084^{***} & -0.009 \\
& (0.024) & (0.344) & (0.046) & (0.017) & (0.119) \\
\beta_{45} & -0.031 & 0.018^{***} & -0.095 & -0.042 & -0.314 \\
& (0.027) & (0.002) & (0.083) & (0.027) & (0.281) \\
\end{array}
\]

Zone-specific variables

\[
\begin{array}{cccccc}
\delta_1 & -0.233 & 0.334 & -0.304^{***} & -0.248^{**} & 0.241 \\
& (0.157) & (0.232) & (0.232) & (0.119) & (0.158) \\
\delta_2 & -0.031^{***} & 0.007 & 0.048 & -0.217^{***} & -0.362^{**} \\
& (0.008) & (0.005) & (0.036) & (0.019) & (0.159) \\
\delta_3 & -0.053 & 0.119^{***} & -0.012 & -0.019 & -0.011 \\
& (0.153) & (0.033) & (0.023) & (0.053) & (0.023) \\
\delta_4 & -0.119^{***} & -0.278 & -0.313^{**} & -0.249^{**} & -0.553^{***} \\
& (0.033) & (0.176) & (0.133) & (0.111) & (0.143) \\
\delta_5 & 0.073 & -0.589^{***} & 0.144 & -0.246^{**} & -1.669 \\
& (0.147) & (0.218) & (0.133) & (0.118) & (1.474) \\
\delta_6 & -0.452^{**} & -0.672^{***} & -0.691^{***} & -0.251^{***} & -0.318^{***} \\
& (0.196) & (0.211) & (0.233) & (0.117) & (0.141) \\
\end{array}
\]
Variance and other model statistics

|                  | Sigma-squared | Gamma       |
|------------------|---------------|-------------|
|                  | 0.551***      | 0.871***    |
|                  | (0.117)       | (0.161)     |
|                  | 0.221**       | 0.663***    |
|                  | (0.097)       | (0.161)     |
|                  | 0.197**       | 0.781***    |
|                  | (0.079)       | (0.161)     |
|                  | 0.256***      | 0.774***    |
|                  | (0.099)       | (0.126)     |

Note: Robust standard errors appear in parentheses. Asterisks denote significance at the following levels. ***=1%, **=5%, *=10%.

In Table 3, the column titled “pooled” denotes that the model is estimated by pooling all the agro-ecological zones in the same regression. While the signs of the estimated meta-frontier are consistent with the pooled frontier, there are differences in the magnitude of the estimated elasticities. In Table 3, in general, the agro-ecological-specific frontier results fit reasonably well as most inputs have their coefficients statistically significant. Notably, there is a lot of variation in magnitudes of parameter estimates across the agro-ecological zones. Some inputs are statistically significant in some regions while others are not, which indicates that input use varies across agro-ecological zones. Because individual parameter estimates of the translog production function are not readily interpretable, we computed output elasticities of inputs and returns-to-scale as shown in Table 4. These elasticities are evaluated at sample means of inputs while returns-to-scale are computed as the sum of the output elasticities. The output elasticities indicate a divergent pattern across agro-ecological zones. We find increasing returns to scale of milk production in the three agro-ecological zones implying that SDFs operate in their first stage of the classical production function and they are still likely to attain their optimal capacity. For policy, these results suggest that SDFs should reduce their average long-term costs by expanding their production scale.

The estimates of agro-ecological zone-specific variables are reported at the bottom of the panel of Table 3. A zones-specific variable with a positive (negative) coefficient implies that the variable has a negative (positive) effect on technical efficiency. In general, most of the variables included under agro-ecological zones-specific variables have a positive effect on technical efficiency across regions (e.g., education, age of household head in years, access to credit, off-farm work participation, group membership and access to extension service by the SDF) which was expected a priori.

### Table 4
Output elasticity estimates and returns to scale

| Output elasticities | LH     | UM     | LM     |
|---------------------|--------|--------|--------|
| \( \beta_1 \)       | 0.034  | 0.183  | 0.258  |
| \( \beta_2 \)       | 0.297  | 0.218  | 0.233  |
| \( \beta_3 \)       | 0.383  | 0.377  | 0.178  |
| \( \beta_4 \)       | 0.114  | 0.321  | 0.219  |
| \( \beta_5 \)       | 0.275  | 0.236  | 0.196  |
| Return to scale     | 1.103  | 1.335  | 1.084  |

### Technical and meta-frontier efficiencies

Table 5 reports the values of SDF efficiency relative to zonal stochastic frontier (TEs), TEs with respect to the meta-frontier and zonal TGRs respectively. TEs are dispersed across the three agro-ecological zones with the LH zone having the highest mean (0.442). The mean agro-ecological zones-specific TE for UM and LM zones are 0.308 and 0.275, respectively. The agro-ecological zonal-specific TEs imply that SDFs are producing below their potential output given the technology available and the
environmental conditions within the respective regions. Thus LH, UM and LM agro-ecological zones have the possibility of improving their respective outputs by 55.8%, 69.2% and 72.5%, respectively, through the adoption of best practices in milk production. The mean technical efficiency across all regions is estimated at 0.355.

As expected, technical efficiency estimates are lower and more dispersed in the meta-frontier model. SDFs in the LH agro-ecological zone achieved the highest mean technical efficiency (0.329), with minimum variation (SD = 0.11), and those in the LM agro-ecological zone had the lowest mean technical efficiency of 0.203. Maximum variation was recorded in the UM agro-ecological zone (SD = 0.194). The mean technical efficiency across all zones is estimated at 0.272. These results suggest that SDFs in UM and LM agro-ecological zones could catch up with their counterparts in the LH agro-ecological zone by adopting the most improved milk production techniques.

**Table 5**

| Model | Item | Agro-ecological zones |
|-------|------|------------------------|
|       |      | LH  | UM  | LM  | Total |
| Zonal-specific TEs | Mean | 0.442 | 0.308 | 0.275 | 0.355 |
|      | Min  | 0.291 | 0.151 | 0.101 | 0.101 |
|      | Max  | 0.647 | 0.768 | 0.672 | 0.768 |
|      | SD   | 0.059 | 0.128 | 0.189 | 0.151 |
| TEs with respect to metafrontier | Mean | 0.329 | 0.259 | 0.203 | 0.272 |
|      | Min  | 0.117 | 0.271 | 0.137 | 0.117 |
|      | Max  | 0.715 | 0.695 | 0.643 | 0.715 |
|      | SD   | 0.104 | 0.194 | 0.174 | 0.137 |
| Regional TGRs | Mean | 0.595 | 0.376 | 0.307 | 0.429 |
|      | Min  | 0.119 | 0.216 | 0.083 | 0.083 |
|      | Max  | 0.748 | 0.816 | 0.688 | 0.816 |
|      | SD   | 0.093 | 0.077 | 0.117 | 0.113 |
| Kruskal-Wallis Test | Chi-squared = 161.016 | P-value = 0.0001 |

The frontier and meta-frontier production estimates for each technology may also be used to calculate the technology gap ratios (TGR) by using Equation (6). TGR measures the proximity of the group-k frontier to the meta-frontier, which represents the current state of knowledge. According to Equation (6), an increase in the TGR implies a decrease in the gap between the group frontier and meta-frontier. Generally, a larger (smaller) TGR implies a higher (lower) gap regarding productive technology between regional-specific frontier and meta-frontier. We adopted the Kruskal-Wallis non-parametric test to examine further whether the TGRs are distinct among groups. The result presented in Table 5 strongly rejects the null hypothesis that there are no differences in the TGRs among different groups. Table 4 shows that the estimates of the mean values of TGR vary more widely than the mean technical efficiency estimates in the meta-frontier model. A TGR value of one (1) implies agro-ecological zonal-specific frontier is tangent to the meta-frontier and there is virtually no room for improvement. The average TGRs for LH, UM and LM agro-ecological zones are 0.595, 0.376 and 0.307, respectively. The results suggest that to be fully technically efficient, SDFs in LH, UM and LM agro-ecological zones ought to close their technology by 40.5%, 62.4% and 69.3%, respectively. Estimated TGR averages 0.429 across all agro-ecological zones. Adopting productivity-enhancing farm technologies from other regions to local conditions could push TEs towards the meta-frontier (Danso-Abbeam & Baiyegunhi, 2014).
The results indicate such adoption and adaptation is highly possible, even LM agro-ecological zone that is relatively far from the meta-frontier.

4. Conclusion and policy implications

This study examines the technical efficiency and technology gap of the three main milk-producing agro-ecological zones, namely, LH, UM and LM agro-ecological zones of Kenya. We controlled unobserved heterogeneity and technological regional differences using Greene (2005a) and metafrontier models, respectively. The empirical analysis is based on comprehensive three-wave household level panel data from across three agro-ecological zones in Kenya. Our results show that there are agro-ecological zones differences differences in technical efficiency in Kenya. The estimated average technical efficiency is 0.44 for the LH zone, 0.31 for UM zone and 0.28 for LM zone. The results suggest that SDFs in all agro-ecological zones suboptimally use available technology in the area, given the regional technology. It is important to note that by accounting for potential technological across agro-ecological zones, we find a disparity between agro-ecological efficiencies and the metafrontier.

We find that there exists a significant technological gap in dairy production across the three agro-ecological zones in Kenya. TGR is different for the three dairy-producing agro-ecological zones of Kenya. The average TGRs for LH, UM and LM agro-ecological zones are 0.595, 0.376 and 0.307, respectively. The results suggest that to be fully technically efficient, SDFs in LH, UM and LM agro-ecological zones ought to close their technology by 40.5%, 62.4% and 69.3%, respectively, which suggests that SDFs in different agro-ecological zones use different production technologies according to the resource endowments and environmental situation of the agro-ecological zone. Consequently, intervention to improve the dairy sector demands these technology differences in the regions be considered.

These results are important for policymakers who wish to increase technical efficiency in dairy production. The result indicate the need to implement agro-ecological zone-specific policies because different enviromnetal conditions characterize each region. SDFs in the LH are near to the frontier and, thus, additional in new technology development is required for relative improvements in smallholder dairy production performance. This intervention helps SDFs in LH agro-ecological zone to use new technology to shift the production frontier upwards and improve milk production performance in the LH agro-ecological zone. Policymakers could minimize the gap for SDFs within UM and LM agro-ecological zones, which allows the inefficient SDFs to learn from the best-performing farms. Moreover, technology adoption and information transfer from the LH agro-ecological zone is crucial to reduce the gap. This policy intervention helps farmers achieve the highest possible output on the metafrontier with the current technology available in the agricultural sector.

The results are based on three agro-ecological zones in Kenya due to data limitations and may not be generalizable to the whole country. Future studies may extend our analysis by including other agro-ecological zones when data becomes available to examine the performance and regional technology gaps of Kenyan dairy-producing farms located in different agro-ecological zones.

Author Declarations
Ethics approval and consent to participate
Not applicable
Consent for publication
Not applicable
**Availability of data and materials**
The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Competing interests**
The authors declare no conflict of interests.

**Funding**
Funding for this effort was by the German Academic Exchange Service (DAAD).

**Author's contributions**
All authors contributed to writing and approved this manuscript to be published.

**Acknowledgements**
The authors are grateful to smallholder dairy farmers in the main milk-producing agro-ecological zones for their overall cooperation in data collection.

**References**

Alem, H., Lien, G., Hardaker, J. B., & Guttormsen, A. (2019). Regional differences in technical efficiency and technological gap of Norwegian dairy farms: a stochastic meta-frontier model. *Applied Economics, 51*(4), 409-421.

Bateki, C., van Dijk, S., Wilkes, A., Dickhoefer, U., & White, R. (2020). Meta-analysis of the effects of on-farm management strategies on milk yields of dairy cattle on smallholder farms in the Tropics. *animal, 14*(12), 2619-2627.

Battese, G. E., & Rao, D. P. (2002). Technology gap, efficiency, and a stochastic metafrontier function. *International Journal of Business and Economics, 1*(2), 87.

Battese, G. E., Rao, D. P., & O’donnell, C. J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of productivity analysis, 21*(1), 91-103.

Danso-Abbeam, G., & Baiyegunhi, L. J. (2020). Technical efficiency and technology gap in Ghana’s cocoa industry: accounting for farm heterogeneity. *Applied Economics, 52*(1), 100-112.

Faostat, F. (2021). Agriculture organization of the United Nations statistics division 2020. Production Available in: [http://faostat3.fao.org/browse/Q/QC/S](http://faostat3.fao.org/browse/Q/QC/S) [Review date: June 2021].

Gelan, A., & Muriithi, B. W. (2012). Measuring and explaining technical efficiency of dairy farms: a case study of smallholder farms in East Africa. *Agrekon, 51*(2), 53-74.

Girma, H. (2019). Estimation of technical efficiency of dairy farms in central zone of Tigray National Regional State. *Heliyon, 5*(3), e01322.

Greene, W. (2005a). Fixed and random effects in stochastic frontier models. *Journal of productivity analysis, 23*(1), 7-32.

Greene, W. (2005b). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of econometrics, 126*(2), 269-303.

KDB. (2018). Kenya Dairy Board Annual Report and Financial Statements for the Year Ended June 30, 2018.

KDB. (2021). Annual report and financial statements for the year ended 30 June 2020.

Ma, W., Bicknell, K., & Renwick, A. (2019). Feed use intensification and technical efficiency of dairy farms in New Zealand. *Australian Journal of Agricultural and Resource Economics, 63*(1), 20-38.

Maina, F., Mburu, J., Gitau, G., & VanLeeuwen, J. (2020). Factors influencing economic efficiency of milk production among small-scale dairy farms in Mukurweini, Nyeri County, Kenya. *Tropical animal health and production, 52*(2), 533-539.
Makau, D., VanLeeuwen, J., Gitau, G., McKenna, S., Walton, C., Muraya, J., & Wichtel, J. (2020). Effects of Calliandra and Sesbania on daily milk production in dairy cows on commercial smallholder farms in Kenya. *Veterinary medicine international, 2020.*

Mareth, T., Scavarda, L. F., Thomé, A. M. T., Oliveira, F. L. C., & Alves, T. W. (2019). Analysing the determinants of technical efficiency of dairy farms in Brazil. *International Journal of Productivity and Performance Management.*

Mareth, T., Thomé, A. M. T., Scavarda, L. F., & Oliveira, F. L. C. (2017). Technical efficiency in dairy farms: research framework, literature classification and research agenda. *International Journal of Productivity and Performance Management, 66*(3), 380-404.

McDermott, J. J., Staal, S. J., Freeman, H., Herrero, M., & Van de Steeg, J. (2010). Sustaining intensification of smallholder livestock systems in the tropics. *Livestock science, 130*(1-3), 95-109.

Mehmood, Y., Rong, K., Bashir, M. K., & Arshad, M. (2018). Does partial quantity rationing of credit affect the technical efficiency of dairy farmers in Punjab, Pakistan? *British Food Journal.*

Muunda, E., Mtimet, N., Schneider, F., Wanyoike, F., Dominguez-Salas, P., & Alonso, S. (2021). Could the new dairy policy affect milk allocation to infants in Kenya? A best-worst scaling approach. *Food policy, 102043.*

O'Donnell, C. J., Rao, D. P., & Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics, 34*(2), 231-255.

Sabasi, D., Shumway, C. R., & Astill, G. M. (2019). Off-farm work and technical efficiency on US dairies. *Agricultural Economics, 50*(4), 379-393.

Skevas, I., Emvalomatis, G., & Brümmer, B. (2018). Heterogeneity of long-run technical efficiency of German dairy farms: A Bayesian approach. *Journal of Agricultural Economics, 69*(1), 58-75.

Wilkes, A., Wassie, S., Odhong, C., Fraval, S., & van Dijk, S. (2020). Variation in the carbon footprint of milk production on smallholder dairy farms in central Kenya. *Journal of Cleaner Production, 121780.*