Journal of Rural and Regional Innovation Studies

E-ISSN: XXXX-XXXX | P-ISSN: XXXX-XXXXX
JORRIS, Vol. 1 (1), 2025 (16-28)
https://journal.unesa.ac.id/index.php/jorris

What Drives Milk Output In Indonesian Smallholder Dairy Farms? Insights From Pagerjurang

Haryo Bimo Budi Indrasto[1*], Nur Hidayah[2], Muslim Affandi [3]

[1]Faculty of Economic and Bussiness, Universitas Muhammadiyah Surakarta, Indonesia
[2] Faculty of Economic and Bussiness, Universitas Muhammadiyah Surakarta, Indonesia
[3]Lembaga Pengembangan Teknologi Pedesaan, Indonesia

*e-mail: b300220233@student.ums.ac.id

Abstract

Low dairy cow milk productivity remains a key constraint to improving farmer welfare and food security in many developing countries, including Indonesia. While prior studies have analyzed dairy productivity, few have explored how both physical and managerial inputs interact under smallholder conditions using econometric approaches that address endogeneity. This study fills that gap by applying a Two-Stage Least Squares (2SLS) and Quantile Regression framework to identify causal and distributional determinants of milk output. Using data from 80 purposively selected respondents, the results reveal that herd size and concentrate feed significantly enhance milk output, while tofu dregs and milking time show no significant impact, implying inconsistencies in feed quality and labor technique. The findings emphasize the importance of feed quality, herd expansion, and scalable managerial interventions to sustainably improve smallholder productivity and local economic resilience in rural Indonesia.

***Keywords*:**
Dairy productivity; Two-Stage Least Squares; Smallholder Farmers

***To cite*:**
Indrasto, HB., Hidayah, N., Affandi, M. (2026). What Drives Milk Output In Indonesian Smallholder Dairy Farms? Insights From Pagerjurang. *Journal of Rural and Regional Studies*, 1(1),x-x,

Introduction

Milk production plays a vital role in Indonesia’s livestock agribusiness, particularly in meeting national food security objectives through the provision of animal-based protein. Despite rising consumer demand for dairy products, domestic milk output remains persistently low, leading to a structural reliance on imports (Williams et al., 2021). Beyond the national level, local production systems often face chronic inefficiencies stemming from poor infrastructure, inconsistent input use, and limited management capacity (Deddefo et al., 2023). Pagerjurang Village exemplifies these challenges; while it has a long-standing tradition of smallholder dairy farming, milk productivity remains stagnant and uneven. A recent study by Hidayah & Perwithosuci (2024) found that smallholder systems often depend heavily on stable partnerships and technical guidance, yet still suffer from environmental and managerial constraints. This setting provides a compelling case for investigating local-level production determinants that, if addressed, could contribute to national self-sufficiency.

Improving dairy productivity requires a deeper understanding of how farm-level inputs such as herd size, barn conditions, milking time, and feed types collectively influence output. While previous studies have highlighted the importance of technical efficiency and nutritional management in improving outcomes (Fiorillo & Amico, 2024), many rely on linear models that fail to address causal inference or variation across farmer types. Moreover, studies focusing on under-researched rural contexts such as Pagerjurang remain scarce, even though such areas may offer unique insights into smallholder resilience and efficiency. This study seeks to fill that gap by employing an econometric approach that accounts for both endogeneity and heterogeneity of response. By focusing on the operational dynamics of smallholders, the findings aim to inform targeted policies that support both productivity and sustainability (Fadilla et al., 2022).

In microeconomic theory, production is viewed as the relationship between inputs and outputs, typically modeled using a production function. The Cobb-Douglas function remains one of the most widely used due to its simplicity, analytical clarity, and ability to estimate both input elasticities and returns to scale (Senga Kiessé et al., 2022). Its log-linear form allows for straightforward interpretation of each input's relative contribution to output (Gujarati, 2009), making it especially relevant in agricultural and livestock contexts. In dairy farming, factors such as the number of lactating cows, time spent milking, barn space, and the use of supplemental feeds particularly concentrates and tofu dregs are considered key inputs influencing milk yield (Setianto N et al., 2025). Differences in how these inputs are combined across farms often lead to significant variations in productivity. Therefore, Cobb-Douglas provides a suitable framework for analyzing smallholder milk production efficiency in rural livestock systems.

In the Indonesian context, rural dairy innovation has increasingly combined technological and institutional approaches. Examples include cooperative-based milk collection centers, feed block technology, and digital herd-recording systems implemented by farmer groups in Lembang and Boyolali. These initiatives exemplify how local innovation complements production theory by improving information flow and efficiency. Moreover, the Cobb–Douglas model assumes constant elasticity of substitution among inputs, allowing efficiency comparisons that are particularly valuable for heterogeneous smallholder systems with mixed resource endowments. A growing body of empirical research supports the use of the Cobb-Douglas production function in analyzing milk output within livestock systems. For example, Bang et al. (2022) found that both the number of lactating cows and the quality of feed significantly impacted daily milk production in Ethiopian smallholder farms. Similarly, Rahmawati et al. (2022) highlighted the importance of production technology and the use of supplemental feed such as concentrates in enhancing dairy productivity. Complementing these findings, Firman et al. (2025) showed that targeted farmer training and improved feeding strategies also played a critical role in increasing output levels. Collectively, these studies demonstrate that productivity in small-scale dairy farming is shaped by a complex interplay of physical, managerial, and technological factors. They further reinforce the relevance of the Cobb-Douglas model as a robust analytical tool to capture such multifactor relationships in smallholder livestock production contexts.

Despite the usefulness of previous studies, many rely solely on linear regression techniques without addressing potential endogeneity issues, especially regarding key variables such as the number of cows. In smallholder dairy systems, decisions to increase herd size and resulting milk output are often interdependent, leading to simultaneity bias when estimated using Ordinary Least Squares (OLS) (Wooldridge et al., 2016). This may produce biased or inconsistent estimates, particularly in contexts where production decisions are influenced by unobserved factors like farmer motivation or access to credit. Furthermore, the effects of input variables are unlikely to be uniform across all producers. Farmers with different herd sizes, labor capacity, or feed access may experience differing marginal returns from the same inputs. Such heterogeneous effects are not captured by mean-based models like OLS, suggesting the need for alternative methods that can account for both causality and distributional differences in production responses among diverse smallholder groups.

In Indonesia, several studies have explored the effects of alternative feed sources such as tofu dregs on dairy productivity Nugroho et al. (2025). However, few have assessed the combined influence of both physical inputs (e.g., herd size, barn space) and managerial practices (e.g., feeding strategies, labor effort) using econometric approaches that address methodological challenges like endogeneity and estimation robustness. Most existing research adopts standard linear models, which may not fully capture the simultaneous and interactive nature of production decisions in smallholder settings. Consequently, the resulting estimates are often partial, overlooking the complex decision-making processes that farmers face when allocating limited resources. This gap highlights the need for more rigorous empirical approaches that can provide unbiased and context-sensitive insights into dairy productivity drivers, particularly in under-researched rural areas where economic, environmental, and social constraints are closely intertwined.

A notable gap in the literature exists at the geographic level, particularly in underrepresented regions such as Pagerjurang. Although the village has a long-established tradition of smallholder dairy farming, it continues to face structural challenges including limited land availability, inconsistent access to quality feed, and diverse management practices that hinder optimal milk productivity. Despite this, no empirical studies have thoroughly examined how the combined effects of key inputs such as herd size, milking time, barn area, and the use of two distinct types of supplemental feed (concentrates and tofu dregs) influence milk output in this specific context. Understanding how these input combinations operate within local conditions is essential for informing policy and designing interventions that are both evidence-based and contextually relevant.

Addressing this gap is crucial to developing more effective strategies for enhancing productivity and ensuring the sustainability of smallholder dairy systems in regions with unique agro-ecological and socio-economic characteristics (Deen et al., 2019). Despite the growing literature on dairy productivity, little empirical work has addressed both causal and distributional dimensions of smallholder production behavior in Indonesia using econometric techniques that correct for endogeneity. This study fills that gap by combining 2SLS and quantile regression to capture the causal relationships and heterogeneous effects of key inputs on milk output in a rural smallholder setting. The analysis contributes to the broader understanding of how farm-level innovation and management practices can enhance local economic resilience.

Methods

This study is designed to identify the key determinants of milk production among smallholder dairy farmers in Pagerjurang Village. To achieve this objective, a quantitative approach is employed using an econometric analysis based on the Cobb-Douglas production function. This method allows for the estimation of input elasticities and returns to scale while addressing potential endogeneity and heterogeneity in production behavior. This chapter presents the research design, data sources, sampling technique, and the analytical framework, including the use of Ordinary Least Squares (OLS), Two-Stage Least Squares (2SLS), and quantile regression.

2.1. *Research Design and Data Collection*

This study adopts a quantitative research design with an econometric approach to analyze the determinants of milk production in smallholder dairy farms. The research was conducted in Pagerjurang Village, a region with a longstanding tradition of dairy farming but persistent productivity challenges. Primary data were collected through structured interviews and direct observation involving 80 dairy farmers, selected using purposive sampling. The selection criteria focused on farmers actively engaged in daily milk production with relatively stable operational records. The survey instrument captured information on production levels, herd characteristics, labor allocation, barn conditions, and feed practices.

2.2. *Theoretical Framework and Model Specification*

The analysis is based on the Cobb-Douglas production function, which models output as a function of multiple inputs using a log-linear form. This functional form is commonly used in agricultural economics due to its ability to estimate input elasticities and returns to scale (Gujarati, 2009). The empirical model is expressed as:

$$\ln MILK_i = \alpha + \beta_1 Cow_i + \beta_2 BARN_i + \beta_3 LABOUR_i + \beta_4 PALET_i + \beta_5 TOFU_i + \varepsilon_i$$

Where:

- $MILK_i$= Daily milk production (liters)
- COW_i = number of lactating cows
- $BARN_i$ = barn area (m^2)
- $LABOUR_i$ = duration of milking per day (minutes)
- $PALET_i$, $TOFU_i$ = dummy variables for use of concentrate and tofu dregs, respectively
- ϵ_i = error term

2.3. *Estimation Strategy*

The estimation procedure began with the application of Ordinary Least Squares (OLS) to establish a baseline relationship between input variables such as herd size, barn area, milking duration, and feed types and daily milk production. This initial model served to identify general patterns of association between farm characteristics and productivity outcomes. To ensure that the OLS estimates were statistically sound, a series of classical assumption tests were conducted. Multicollinearity was assessed using Variance Inflation Factors (VIF), confirming that no severe interdependence existed among the explanatory variables (Wooldridge et al., 2016). Heteroskedasticity was examined using the Breusch-Pagan test, and the normality of residuals was also evaluated to confirm distributional assumptions. These diagnostic checks were necessary to validate the consistency and interpretability of the OLS model before proceeding to more advanced estimation methods. The results of these tests provided insights into potential model limitations and helped identify the presence of specification issues, particularly related to the possibility of endogeneity among key variables such as herd size.

Given the potential endogeneity of herd size where the decision to maintain more cows could be influenced by previous or expected levels of milk production a Two-Stage Least Squares (2SLS) estimation was conducted. This method is particularly useful for addressing simultaneity bias, which can arise when the explanatory variable is jointly determined with the outcome. In the first stage, barn area was used as an instrumental variable (IV) for herd size, based on the logical premise that farmers with larger barns are more likely to keep more cows, while barn size itself is not directly influenced by short-term productivity shocks (Gujarati, 2009). The validity of the instrument was tested through the significance of the first-stage regression results, ensuring that the instrument was both relevant and sufficiently correlated with the endogenous regressor. By employing the 2SLS approach, the model was better equipped to provide causal interpretations of the relationship between input variables and milk production, offering a more rigorous alternative to OLS in the presence of endogenous regressors.

2.4. Robustness and Distributional Analysis

To address the potential endogeneity of the number of cows, the study employs a Two-Stage Least Squares (2SLS) estimation using barn area as an instrumental variable. The choice of barn area is based on the assumption that it correlates with the decision to maintain more cows yet remains uncorrelated with unobserved shocks to milk productivity a requirement for a valid instrument. This approach mirrors the methodology used by Qi et al. (2022) in China's dairy sector, where spatial variables served as credibly exogenous instruments in a 2SLS estimation. Instrument strength was confirmed via a first-stage regression and a Cragg–Donald F-statistic above conventional thresholds, affirming its suitability.

To assess the robustness of our findings and explore heterogeneous effects across the distribution of milk production, the third phase employed quantile regression at the 25th, 50th, and 75th percentiles. Unlike OLS or 2SLS, which only estimate average effects, quantile regression enables examination of how variable influences vary across performance levels a method increasingly used in agricultural economics. This approach is especially relevant in smallholder systems, where access to inputs, labor quality, or managerial practices can differ markedly. For example, Qi et al. (2022) demonstrated that quantile IV models better capture varying treatment effects than average-based models. All analyses were conducted in Stata 17, with diagnostic procedures including instrument validity tests and heteroskedasticity checks to ensure reliable estimation results.

Result

3.1. Descriptive Statistics

The preliminary analysis presents descriptive statistics of all variables used in the production function model. On average, smallholder dairy farmers in the sample produced approximately 44.58 liters of milk per day, with outputs ranging from 21.9 to 67.5 liters. The number of lactating cows varied from 3 to 9, with a mean of 5.25 cows per household. The average barn area was 22.08 square meters, indicating differences in farm scale and capacity. Farmers reported spending an average of 9.57 minutes per day milking their cows, though this input exhibited relatively low variation. Regarding feed practices, 73.7% of respondents used concentrate feed, while 81% reported utilizing tofu dregs as a supplemental input. These figures suggest diverse combinations of physical and managerial inputs among smallholders. Further detail on the characteristics of each variable is presented in Table 1, which summarizes the mean, standard deviation, and range for each input and output variable.

Table 1. Descriptive Statistic

Variable	Mean	Std. Dev.	Min	Max	N
Milk Production	44.583	12.369	21.9	67.5	80
Herd Size	5.25	1.945	3	9	80
Barn Area	22.082	8.298	10.5	48.75	80
Milking Duration	9.575	1.581	7	13	80
Use of Concentrate Feed	0.737	0.442	0	1	80
Use of Tofu Dregs	0.81	0.402	0	1	80

Source: Author, 2025

The variability observed in key input variables, such as herd size and feeding practices, indicates potential differences in productivity drivers. These patterns provide a rationale for the subsequent econometric estimation to identify significant determinants of milk output.

3.2. Model Validation and Diagnostic Tests

The regression model underwent a series of validity tests to ensure compliance with the classical assumptions. As shown in Table 2, the Shapiro-Wilk test for residual normality yielded a test statistic of 1.326 with a p-value of 0.0924. Since the p-value exceeds the 0.05 threshold, the null hypothesis of normally distributed residuals cannot be rejected, indicating that the residuals follow a normal distribution at the 5% significance level. In contrast, the Breusch-Pagan test for heteroskedasticity produced a test statistic of 14.85 with a p-value of 0.0001, which is well below the 0.05 significance level. This result indicates the presence of heteroskedasticity in the model, suggesting that the error variance is not constant across observations. Finally, the test for multicollinearity based on the Mean Variance Inflation Factor (VIF) showed a value of 5.92, which is below the commonly accepted threshold of 10. This suggests that multicollinearity is not a significant concern among the explanatory variables. Given the evidence of heteroskedasticity, the regression analysis proceeded using robust Ordinary Least Squares (OLS) estimation to produce consistent and reliable coefficient estimates.

Table 2. Validity Test

Validity Test	Stat	P-Value
Normality (Saphiro-Wilk)	1,326	0,0924
Heteroskedasticity (Breusch-Pagan)	14,85	0,0001
Multicollinearity VIF	Mean VIF	
	5.92	

Source: Author, 2025

Based on the validity tests in Table 2, the presence of heteroskedasticity justifies the use of robust standard errors. Therefore, the subsequent regression analysis begins with robust OLS estimation.

3.3. OLS Estimation Results

The results of the robust OLS regression are presented in Table 3. Several input variables exhibit statistically significant effects on milk production, particularly those related to herd size, housing, and feed practices. Meanwhile, a few variables show limited or insignificant influence, reflecting potential variation in farmer practices or contextual factors.

Table 3. Ordinary Least Square (Robust)

Variabel	Coeff	t-stat	P-value
Cow	0.2755	10.461	0.000
Barn	0.0239	4.643	0.000
Labour	0.0008	0.091	0.934
Palet	0.4609	7.370	0.000
Tofu	0.0243	0.633	0.508
Constant	2.4705	19.361	0.000

Source: Author, 2025

Notes: Wald F(5, 64) = 104.29, Prob > F = 0.000, R^2 = 0.7941, N = 70. Robust standard errors used.

The robust Ordinary Least Squares (OLS) estimation reveals that several input variables significantly influence daily milk production among farmers in Pagerjurang shown in Table 2. The number of cows shows a coefficient of 0.2755 and is statistically significant at the 1% level (p-value = 0.000), indicating that a 1% increase in the number of lactating cows leads to a 0.2755% rise in milk output, holding other variables constant. This finding underscores the central role of herd size in smallholder dairy systems. Additionally, barn area is also significant (coefficient = 0.0239, p-value = 0.000), suggesting that physical space positively affects milk productivity. Larger barn space may enhance cow comfort and health, which, in turn, contributes to better milk yields. These results support the notion that both biological and physical capital are critical to improving dairy production efficiency in rural contexts.

Another important determinant is the use of concentrate feed (palet), which has a strong positive and statistically significant coefficient of 0.4609 (p-value = 0.000). This implies that farmers who use concentrate feed can expect approximately 46% higher milk output compared to those who do not, all else equal. The magnitude of this effect highlights the importance of feed quality in enhancing productivity without increasing herd size. Concentrates are known to provide high nutritional value essential for lactation. In contrast, the use of tofu dregs shows a positive but statistically insignificant coefficient of 0.0243 (p-value = 0.508), suggesting that although tofu dregs are commonly used as a low-cost feed alternative, their contribution to milk production remains inconclusive in this model. This may be due to variability in nutritional content or feeding consistency among households.

The labor variable yields a very small and statistically insignificant coefficient of 0.0008 (p-value = 0.934), indicating that variation in milking time does not significantly impact milk output. This could imply that labor quality or technique matters more than duration. Overall, the robust OLS model yields an R^2 of 0.7941, indicating that approximately 79.41% of the variation in milk production is explained by the included variables. The Wald F-statistic of 104.29 with a p-value of 0.000 confirms that the model is statistically significant as a whole. These results provide strong empirical support for identifying the most influential inputs in smallholder milk production systems and offer policy-relevant insights for improving efficiency in rural dairy farming.

The non-significant effect of tofu dregs may be attributed to inconsistent nutritional quality, since tofu residues vary in protein and moisture content depending on processing and storage. This variability can dilute feed efficiency and obscure its contribution to milk yield. Similarly, milking duration shows no significant influence, suggesting that productivity depends more on milking skill and timing consistency than on duration. These findings highlight that feed quality management and technical training may be more important than simply extending work time or using low-value feed substitutes. Improved farmer education on feed composition, nutrient ratios, and milking hygiene could therefore yield greater productivity gains than physical expansion alone.

3.4. 2SLS Estimation Results

To address potential endogeneity in the OLS estimation, particularly regarding the number of cows, a Two-Stage Least Squares (2SLS) model was employed using barn area as an instrumental variable. Compared to the OLS results, the 2SLS model produced a lower coefficient for the cow variable (0.1455 vs. 0.2755), suggesting that the initial OLS estimation may have overstated the true impact due to simultaneity bias. Despite this adjustment, the effect of cow ownership on milk production remains positive and statistically significant, reinforcing the role of herd size as a primary driver of output. The instrument used passed the underidentification and weak instrument tests, with a Kleibergen-Paap chi-square of 16.91 ($p < 0.001$) and a Cragg-Donald F-statistic of 502.23, confirming its strength and validity. These results highlight the importance of addressing endogeneity in productivity analysis to ensure the robustness and reliability of policy-relevant estimates.

Table 4. Two Stage Least Square

Variabel	Coeff	z-stat	P-value
Cow	0.1455	16.08	0.000
Labour	0.0027	0.23	0.815
Palet	0.2036	5.98	0.000
Tofu	0.0192	-0.51	0.610
Constant	2.8303	20.01	0.000
Observation	80	**Weak IV F**	502,226
R^2	0,7941	**OverID p-value**	0,000
Underidentification (Kleibergen-Paap): $\chi^2 = 16.91$, *p-value* = 0.000			

Source: Author, 2025

Note: The Barn variable is used as an instrument for the Cow variable and is not included as a direct control variable. The estimation was conducted using robust standard errors.

As presented in Table 3, the Two-Stage Least Squares (2SLS) estimation confirms that the number of cows has a statistically significant positive effect on milk production, even after addressing endogeneity. The estimated coefficient for the Cow variable is 0.1455 with a p-value of 0.000, implying that a 1% increase in herd size leads to a

0.15% increase in milk output, ceteris paribus. This result is consistent with expectations but notably lower than the OLS estimate of 0.2755. The difference suggests that the OLS model may have overstated the effect due to simultaneity bias farms with higher productivity are more likely to maintain larger herds. The 2SLS approach thus offers a more conservative and likely more accurate estimate of the causal effect of herd size.

In addition to the herd size, Table 3 shows that the use of concentrate feed (Palet) remains a strong determinant of milk productivity. The coefficient is 0.2036 with a p-value of 0.000, reaffirming the positive impact of nutrient-rich feed on output. This result mirrors the OLS findings and is supported by prior studies on dairy nutrition. On the other hand, the variables Labour (milking time) and Tofu (use of tofu dregs) exhibit no statistically significant effects, with p-values of 0.457 and 0.542, respectively. These findings suggest that variation in labor duration and the use of tofu dregs are not consistent predictors of milk output, possibly due to differences in technique, intensity, or feed quality not captured by binary variables. Their insignificance strengthens the validity of the model by highlighting the inputs that truly drive output.

The strength and validity of the instrumental variable are also supported by the diagnostic statistics reported in Table 3. The Cragg-Donald F-statistic equals 502.226, far exceeding the critical threshold for weak instruments, indicating that barn area is a strong and relevant instrument for cow ownership. The Kleibergen-Paap underidentification test yields a chi-square value of 16.91 with a p-value of 0.000, confirming the model is properly identified. Additionally, the Hansen J test indicates that the instrument satisfies the exclusion restriction, as no evidence of overidentification is present. These results collectively support the reliability of the 2SLS estimates. Taken together, the findings validate that both herd size and concentrate feed are key inputs in explaining milk production, while the econometric strategy provides robust causal inference in the presence of endogeneity.

3.5. Robustness Check Using Quantile Regression

Quantile regression was employed to explore how the effects of input variables differ across various levels of milk production. This approach allows for the detection of heterogeneous effects that may not be captured by mean-based estimations. Table 4 presents the estimated coefficients for the 25th, 50th, and 75th percentiles of milk output.

Table 5. Multiple Quantile Regression

Variabel	Q25 Coeff (SE)	Q50 Coeff (SE)	Q75 Coeff (SE)
Cow	0.26177*	0.2461*	0.2709*
Barn	0.0224*	0.0189*	0.0254*

Labour	0.0008	0.0012	0.0026
Palet	0.4295*	0.3859*	0.4887*
Tofu	0.0276	0.0119	0.0069
Constant	2.4979*	2.5875*	2.6425*
Observation	80	80	80
Pseudo R^2	0.6295	0.6642	0.6382

Source: Author, 2025

Quantile regression analysis was employed to capture heterogeneous effects of input variables across different levels of milk production. By examining the 25th, 50th, and 75th percentiles (Q25, Q50, Q75), the model allows for a more nuanced understanding of how production inputs behave among lower-, middle-, and higher-performing dairy farmers. Table 4 presents the estimated coefficients and standard errors for each quantile, revealing clear patterns of consistency and divergence across groups.

Across all three quantiles, the number of cows (Cow) showed a positive and statistically significant effect on milk production, with slightly increasing coefficients from Q50 (0.2461) to Q75 (0.2709). This suggests that herd size contributes to production across all levels, but particularly benefits high-output farmers. Similarly, the barn area (Barn) was statistically significant at all quantiles, though its coefficients remained small, indicating a modest but stable contribution to milk productivity regardless of scale. In contrast, the labour variable (Labour) was not statistically significant in any quantile, implying that differences in milking time do not account for variations in milk output across farmer groups.

Palet (concentrate feed) emerged as the most influential input, with strong positive and significant effects across all quantiles, and its impact rising toward the upper end (from 0.3859 at Q50 to 0.4887 at Q75). This highlights the critical role of high-quality nutrition in boosting output, especially among top-performing farmers. Conversely, the tofu variable (Tofu), representing the use of tofu dregs, was not statistically significant in any quantile, indicating limited effect on productivity. The increasing intercept (Constant) values from Q25 to Q75 reflect baseline productivity differences across farmer groups. The pseudo R^2 values (ranging from 0.6295 to 0.6642) suggest that the quantile regression models have a reasonably good explanatory power across all levels of milk production.

Discussion

The findings from this study indicate that two main factors the number of lactating cows and the use of concentrate feed are the most consistent and significant determinants of milk production in smallholder dairy farms in Pagerjurang Village. These variables exhibited positive and statistically significant effects across all three estimation methods, namely Ordinary Least Squares (OLS), Two-Stage Least Squares

(2SLS), and quantile regression. This consistency supports the importance of herd size and nutritional supplementation in enhancing productivity. Olagunju et al. (2022) similarly found that scale and feed practices are central to improving total factor productivity in dairy systems. The instrumental variable approach using barn area to address the endogeneity of herd size further reinforces the causal interpretation of this relationship. These results suggest that policies focused on herd expansion and improved feeding strategies can generate substantial productivity gains in smallholder contexts with limited resources.

Beyond the core determinants, the quantile regression approach provided additional insights by revealing how input effects differ across productivity levels. Specifically, barn area was found to significantly influence milk production only among higher-performing farmers, as indicated by its significance at the 75th percentile. This implies that while concentrate feed benefits farmers across all productivity levels, infrastructural improvements such as barn space are more effective when the overall management system is already optimized. These results are consistent with Bang et al. (2022), who demonstrated that barn quality and feeding systems significantly influence cow productivity and welfare in small-scale tropical dairy systems. The implication is that a one-size-fits-all policy may not work. Instead, support for low-productivity farmers should focus on essential feed access, while high-productivity farmers could benefit more from structural investments.

Meanwhile, other variables such as milking time and the use of tofu dregs showed no consistent or significant effect in any of the models. This may reflect homogeneity in labor practices or insufficient variation in feed quality, especially for tofu dregs, which are likely used as supplemental rather than primary feed. Although these factors did not appear statistically significant, they still offer important contextual understanding about prevailing practices in Pagerjurang. Their low impact also highlights the need for targeted interventions that focus on inputs with proven effectiveness. The combination of OLS, 2SLS, and quantile regression has shown value in capturing both general patterns and subgroup-specific dynamics. This suggests that policymakers and stakeholders should adopt differentiated strategies depending on farmers' scale, management practices, and productivity tier.

Given the varying significance and magnitude of input effects especially among variables such as herd size and concentrate feed it becomes critical to evaluate whether these relationships hold consistently across estimation strategies. Relying solely on one method, such as OLS, may obscure important nuances, particularly in the presence of endogeneity or heterogeneous responses. Therefore, a methodological comparison is essential to validate the robustness of findings and to capture the complexity of decision-making in smallholder dairy systems. The following section presents a systematic comparison of three econometric approaches

OLS, 2SLS, and quantile regression each offering distinct advantages in addressing structural and distributional aspects of the data.

The OLS model revealed that herd size and concentrate feed significantly influence milk production. However, this approach assumes all explanatory variables are exogenous, which may not hold in real-world settings. For instance, farmers with greater output may be more likely to increase herd size, introducing simultaneity bias (Ullah et al., 2021). To address this concern, 2SLS was used with barn area as an instrument for herd size. The similarity in coefficient direction and significance across both OLS and 2SLS supports the reliability of results, aligning with findings from Qi et al. (2022), who emphasize the importance of proper instrumentation in agricultural production models. Nonetheless, minor changes in coefficient magnitudes suggest that OLS may slightly overestimate the impact of herd size without controlling for endogeneity.

Quantile regression further deepens the analysis by capturing the distributional impacts of production inputs across varying productivity levels. The effect of herd size and concentrate feed appears more pronounced among higher-output farmers, especially at the 75th percentile, suggesting that better-resourced or better-managed farms can leverage these inputs more effectively. This observation echoes the findings of Meiners & Torres (2022), who emphasized productivity heterogeneity across different farm size groups in Brazil, and Bang et al. (2022), who observed similar disparities in Vietnamese dairy systems. Antonioli et al. (2024) further highlighted that the efficiency of input use is closely tied to managerial capacity and peer-learning dynamics. These studies collectively reinforce that a one-size-fits-all policy is unlikely to yield optimal results; instead, differentiated interventions based on farm scale and capability should be prioritized.

The positive and consistent impact of concentrate feed across all model specifications is strongly supported by recent literature. Akzar et al. (2025) demonstrated that the adoption of bundled feed technologies particularly commercial concentrates significantly improved milk output among Indonesian smallholders. Firman et al. (2025) similarly found that concentrate feeding was a key strategy among farmers in Pangalengan to overcome forage shortages due to limited land and seasonal variability. This aligns with the current study's context in Pagerjurang, where farmers may face similar constraints. Such findings underscore that targeted interventions in feeding strategies can produce immediate and measurable improvements in productivity, especially in tropical smallholder systems constrained by limited land and erratic climatic conditions.

Beyond feed and herd size, infrastructure-related factors such as barn conditions and milking practices also deserve attention. While barn area was excluded from the 2SLS model as a direct regressor and instead used as an instrument, its significance in OLS results and theoretical linkage to cow-holding capacity remains relevant. Fadillah et al. (2023) found that better barn hygiene and farmer training were associated with lower bacterial contamination in milk, reflecting both improved quality and management. Mukasafari et al. (2025) also emphasized that barn maintenance and milking frequency substantially influenced milk yield and composition in Rwandan farms. These findings suggest that even infrastructural and managerial elements though sometimes indirectly measured are integral to enhancing dairy farm performance, and should not be overlooked in policy design.

On the other hand, variables such as milking time and the use of tofu dregs were not statistically significant in this study. This outcome is echoed in Nugroho et al. (2025), who showed that labor productivity in smallholder dairy farms is not solely determined by working hours, but by labor quality, training, and operational efficiency factors not always quantifiable in raw time data. Furthermore, tofu dregs, often used as a supplementary feed, may not provide sufficient nutritional value to significantly influence milk output unless paired with more comprehensive feeding strategies. As Olagunju et al. (2020) pointed out, input effects tend to be marginal without the support of effective managerial practices. Therefore, these findings reinforce the need for context-sensitive selection of input strategies and targeted policies that reflect smallholders' varying capabilities and constraints.

The findings of this study offer several practical implications for policy design in smallholder dairy systems, particularly in regions such as Pagerjurang. The results reveal that increasing the number of cows and the use of concentrate feed are the most consistent and significant determinants of milk output. This suggests that policies promoting access to livestock financing and the provision of affordable, high-quality feed can effectively enhance productivity. Furthermore, since barn area indirectly influences production through its association with herd size, farmer training on barn planning and spatial management could also help improve production capacity incrementally and sustainably.

Additionally, the quantile regression results highlight that the impact of inputs varies across productivity levels. This indicates that policy interventions should not adopt a one-size-fits-all approach but instead be tailored to the specific needs of different farmer segments. For low-productivity farmers, policies should focus on improving managerial skills, feed access, and basic infrastructure, while for high-performing farmers, strategies should prioritize technical upgrades, efficiency gains, and enterprise scaling. Similar recommendations were made by Meiners & Torres (2022) in their study on heterogeneity in agricultural productivity. Therefore, differentiated, data-driven interventions are more likely to be effective and contextually relevant. This calls for more granular farmer profiling at the village level to guide program design and resource allocation. By aligning input-specific support with productivity tiers, stakeholders can better optimize outcomes and promote inclusive growth in the dairy sector.

Despite providing meaningful insights, this study has several limitations that should be acknowledged. First, the analysis is based on cross-sectional data collected from a single village, which limits the generalizability of the findings to other regions with different agroecological or institutional conditions. Future research should consider a broader geographic scope or panel data to capture temporal variations and strengthen causal interpretations. Second, although instrumental variable methods were employed to address endogeneity, the validity of instruments may still be debated, especially in contexts with limited data on unobserved productivity shocks.

Another limitation lies in the measurement of certain variables. For instance, labor quality was proxied only by milking time, which may not fully reflect skill differences or task complexity. Similarly, feed quality was represented by binary indicators of concentrate and tofu dreg use, without capturing quantity or nutritional content. These simplifications could mask more nuanced effects. Future studies may benefit from incorporating more granular data, including detailed input quantities, management indicators, and milk quality metrics, to enrich the understanding of productivity drivers in smallholder dairy systems.

From a policy standpoint, the findings highlight that scalable dairy innovation should not only target technical improvements but also social inclusion. Women perform the majority of milking and feed preparation tasks in many smallholder households, yet often lack access to credit and training programs (Nugroho et al., 2025). Integrating gender-responsive extension services and youth-oriented dairy entrepreneurship initiatives such as village-based cooperative startups can strengthen innovation adoption and generational continuity. Such efforts are increasingly relevant given the rising participation of young people in agri-based innovation across Indonesia's rural regions (Nurhayati et al., 2024). These approaches align with Indonesia's rural transformation agenda and ensure that productivity gains translate into inclusive local economic resilience.

Conclusion

This study examined the key determinants of milk production among smallholder dairy farmers in Pagerjurang using a Cobb–Douglas production framework. Through a combination of OLS, 2SLS, and quantile regression techniques, the analysis revealed that herd size and the use of concentrate feed are the most consistently significant predictors of milk output. The number of cows had a positive and significant effect in both OLS and 2SLS models, even after addressing potential endogeneity using barn area as an instrument. Similarly, the use of concentrate feed showed a robust influence, particularly among higher productivity quantiles. In contrast, other variables such as milking time and tofu dregs showed no consistent or significant impact across models. These findings emphasize that physical scale and nutritional inputs play a central role in determining milk production outcomes in smallholder systems, whereas labor time and supplemental feeds may have less pronounced effects.

Beyond average relationships, the use of quantile regression allowed for a deeper understanding of how input effectiveness varies across farmer performance levels. The results suggest that policies aiming to improve milk production should not be uniform but rather tailored to different segments of farmers based on their current productivity tier. For instance, support programs could prioritize concentrate subsidies or improved breeding strategies for more productive farms, while capacity building and technical assistance may better suit lower-performing groups. Moreover, the findings reinforce the value of combining econometric techniques to address endogeneity and explore heterogeneous effects in agricultural productivity studies. Overall, this research contributes to the empirical literature by offering context-specific insights into input efficiency and by informing the design of evidence-based interventions to enhance the sustainability and profitability of smallholder dairy farming in Indonesia.

Declarations

No ethical issues have arisen during the study and all procedures followed comply with the ethical standards.

Authors contribution statement

Author Haryo Bimo Budi Indrasto was responsible for conceptualization, study design, and data analysis. Author Nur Hidayah conducted data collection, data processing, and drafted the initial manuscript. Author Muslim Affandi contributed to the literature review, validation of results, and manuscript editing. All authors have read and approved the final version of the manuscript.

Funding statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Declaration of interests statement

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

References

Akzar, R., Peralta, A., & Umberger, W. (2025). Adoption of dairy feed technology bundles improves smallholder dairy farmers' milk production. *Journal of Agribusiness in Developing and Emerging Economies*, *15*(2), 269–287. https://doi.org/https://doi.org/10.1108/JADEE-12-2022-0267

Antonioli, F., Severini, S., & Vigani, M. (2024). Foreign labor, peer-networking and agricultural efficiency in the Italian dairy sector. *Agribusiness*. https://doi.org/https://doi.org/10.1002/agr.21970

Bang, N. N., Chanh, N. V, Trach, N. X., Khang, D. N., Hayes, B. J., Gaughan, J. B., & McNeill, D. M. (2022). Multivariate analysis identifying the main factors associated with cow productivity and welfare in tropical smallholder dairy farms in Vietnam. *Tropical Animal Health and Production*, *54*(5), 313. https://doi.org/https://doi.org/10.1007/s11250-022-03303-7

Deddefo, A., Mamo, G., Asfaw, M., & Amenu, K. (2023). Factors affecting the microbiological quality and contamination of farm bulk milk by Staphylococcus aureus in dairy farms in Asella, Ethiopia. *BMC Microbiology*, *23*(1), 65. https://doi.org/https://doi.org/10.1186/s12866-022-02746-0

Deen, A. U., Tyagi, N., Yadav, R. D., Kumar, S., Tyagi, A. K., & Singh, S. K. (2019). Feeding Balanced Ration Can Improve The Productivity And Economics Of Milk Production In Dairy Cattle: A Comprehensive Field Study. *Tropical Animal Health And Production*, *51*, 737–744. https://doi.org/https://doi.org/10.1007/s11250-018-1747-8

Fadilla, M., Nurmawati, E., Fasa, M. I., & Suharto. (2022). Peran Sumber Daya Alam Bagi Pembangunan Berkelanjutan Di Indonesia Dalam Persfektif Islam. *JEKSYAH Islamic Economics Journal*, *2*, 54–63. https://doi.org/https://doi.org/10.54045/jeksyah.v2i01.70

Fadillah, A., Borne, B. H., Poetri, O. N., Hogeveen, H., Slijper, T., Pisestyani, H., & Schukken, Y. H. (2023). Evaluation of factors associated with bulk milk somatic cell count and total plate count in Indonesian smallholder dairy farms. *Frontiers in Veterinary Science*, *10*, 1280264. https://doi.org/https://doi.org/10.3389/fvets.2023.1280264

Fiorillo, V., & Amico, B. M. (2024). Milk quality and economic sustainability in dairy farming: A systematic review of performance indicators. *Dairy*, *5*(3), 384–402. https://doi.org/https://doi.org/10.3390/dairy5030031

Firman, A., Mauludin, M. A., Salsabila, K. N., & Kusmayadi, T. (2025). The Smallholder Dairy Farmer Strategies to Address Feed Scarcity in Pangalengan, Indonesia. *Jurnal Peternakan*, *22*(1), 1–9. https://doi.org/10.24014/jupet.v22i1.26055

Gujarati, D. N. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.

Hidayah, N., & Perwithosuci, W. (2024). Farmers' perception of climate change: A study in Tulung Sub-District, Indonesia. *IOP Conference Series: Earth and Environmental Science*, *1302*(1), 12140. https://doi.org/10.1088/1755-1315/1302/1/012140

Meiners, P. G., & Torres, M. D. O. (2022). Heterogeneity in agricultural factor productivity across and within farm size groups in Brazil. *Revista Brasileira de Economia*, *76*(1), 72–91. https://doi.org/https://doi.org/10.5935/0034-7140.20220005

Mukasafari, M. A., Mpatswenumugabo, J. P., Ndahetuye, J. B., Wredle, E., & Båge, R. (2025). Management factors affecting milk yield, composition, and quality on smallholder dairy farms. *Tropical Animal Health and Production*, *57*(2), 41. https://doi.org/https://doi.org/10.1007/s11250-025-04294-x

Nugroho, Haryati, N., & Fajriyah, K. (2025). Factors Influencing the Labor Productivity in Smallholder Dairy Farms (Case Study in Kediri Regency, East Java, Indonesia. *5th International Conference on Environmentally Sustainable Animal Industry (ICESAI 2024*, 485–491. https://doi.org/10.2991/978-94-6463-670-3_52

Nurhayati, S., Nurjaman, N., & Rukanda, N. (2024). Innovation, Empowerment, And Community Development: Comprehensive Analysis Of Rural Youth-Driven Coffee Production. *Masyarakat Madani: Jurnal Kajian Islam Dan Pengembangan Masyarakat*, *9*(2). https://doi.org/10.24014/jmm.v9i2.32993

Olagunju, K. O., Patton, M., & Feng, S. (2020). Estimating the impact of decoupled payments on farm production in Northern Ireland: An instrumental variable fixed effect approach. *Sustainability*, *12*(8), 3222. https://doi.org/https://doi.org/10.3390/su12083222

Olagunju, K. O., Sherry, E., Samuel, A., & Caskie, P. (2022). Unpacking total factor productivity on dairy farms using empirical evidence. *Agriculture*, *12*(2), 225. https://doi.org/https://doi.org/10.3390/agriculture12020225

Setianto N. A, Subagyo Y, & Wakhidati Y. N. (2025). System Dynamic Modelling for the Development of Smallholder Dairy Farming in Java Island, Indonesia. *Systems Research and Behavioral Science*. https://doi.org/https://doi.org/10.1002/sres.3202

Qi, Y., Han, J., Shadbolt, N. M., & Zhang, Q. (2022). Can the use of digital technology improve the cow milk productivity in large dairy herds? Evidence from China’s Shandong Province. *Frontiers in Sustainable Food Systems*, *6*, 1083906. https://doi.org/https://doi.org/10.3389/fsufs.2022.1083906

Rahmawati, Y., Dwijayanto, A., & Nugroho, W. A. (2022). Pemberdayaan Peternak Sapi Perah Melalui Pelatihan Pembuatan Produk Susu Varian Rasa Di Pulung Ponorogo. *Amalee: Indonesian Journal Of Community Research And Engagement*, *3*(1), 205–214. https://doi.org/https://doi.org/10.37680/amalee.v3i1.1306

Senga Kiessé, T., Heijungs, R., & Corson, M. S. (2022). Modeling Production Efficiency And Greenhouse Gas Objectives As A Function Of Forage Production Of Dairy Farms Using Copula Models. *Environmental Modeling & Assessment*, *27*(3), 413–424. https://doi.org/https://doi.org/10.1007/s10666-022-09831-8

Ullah, S., Zaefarian, G., & Ullah, F. (2021). How to use instrumental variables in addressing endogeneity? A step-by-step procedure for non-specialists. *Industrial Marketing Management*, *96*, 1– 6. https://doi.org/https://doi.org/10.1016/j.indmarman.2020.03.006

Williams, M., Murphy, C. P., Sleator, R. D., Ring, S. C., & Berry, D. P. (2021). Genetic And Nongenetic Factors Associated With Lactation Length In Seasonal-Calving, Pasture-Based Dairy Cows. *Journal Of Dairy Science*, *104*(1), 561–574. https://doi.org/https://doi.org/10.3168/jds.2020-18941

Wooldridge, J. M., Wadud, M., & Lye, J. (2016). *Introductory econometrics: Asia pacific edition with online study tools 12 months*. Cengage AU.