The cost of joining forces: collective organisations and the economic performance of dairy farms in South West England

Authors:

Paolo Nota[1], Daniele Curzi[1*], Mauro Vigani[2,3]

Affiliations:

[1] Department of Environmental Science and Policy, University of Milan, Italy.

[2] European Commission, Directorate-General for Agriculture and Rural Development (DG AGRI), Brussels, Belgium.

[3] Countryside and Community Research Institute, University of Gloucestershire, Cheltenham, UK.

* corresponding author daniele.curzi@unimi.it

Please cite this article as:

Nota P., Cruzi D., Vigani M., (in press). The cost of joining forces: collective organisations and the economic performance of dairy farms in South West England ,*Bio-Basedand Applied Economics*, Just Accepted. DOI:10.36253/bae-17342

Abstract

This note highlights the importance of sale agreements for the economic performance of dairy farms in South West England. The reduction in public intervention in milk supply governance led to market coordination through supply chain agreements (SCAs), resulting in unbalanced power dynamics that reduced farmers' bargaining power and made them vulnerable to unfair practices. Collective organisations are seen as a solution to protect farmers. Using Coarsened Exact Matching, the paper estimates the effect of collective organisation participation on selling prices, production costs, and markup in a sample of 200 farms in Somerset and Devon. The results show that collective

organisations are associated with 7.3% higher production costs and 23 percentage points lower markup, mainly due to costs associated with milk collection, transport, storage, marketing, and adherence to strict quality standards. However, these costs are counterbalanced by secured market access, stronger negotiating power, and better alignment with market demand. These findings underscore the need to further investigate collective governance in developed agricultural markets.
Keywords: dairy farming, collective organisations, markup, coarsened exact matching

JEL codes: Q12, Q13

1. Introduction

This note highlights the importance of sales arrangements for the economic viability of farms, analysing the case of dairy farming in Somerset and Devon, England. With the abolition of milk quotas in 2015, the dairy sector experienced a shift towards a more market-driven model. Moreover, since Brexit the United Kingdom (UK) has moved away from the European Union's Common Agricultural Policy (CAP) and developed a new framework centred on the Environmental Land Management (ELM) schemes. While ELM emphasizes environmental sustainability and productivity, it does not directly address key challenges for small producers, particularly those arising from vertical and horizontal relationships with other companies in the supply chain.

The English dairy sector has particularly experienced the effects of these changes. Since 2015, the number of dairy farms in England declined by about 3.3% every year[1]. Continued pressure on dairy margins contributed to the decline, with growth of milk-to-feed price ratios of only 0.8% per year insufficient to generate sustainable profits[2]. In addition, the abolition of the quota system changed the mechanism for aligning milk production to consumer demand, giving a more important role to supply chain agreements (SCAs). Within this evolving supply chain, smaller UK dairy producers deal directly with larger companies – both upstream (e.g., feed and energy suppliers) and downstream (e.g., dairy processors and retailers) – which have greater bargaining power and often set their private quality standards (Sexton, 2012). The growing gap in access to market information between farmers and their trading partners can also lead to unfair trading practices, further disadvantaging smaller producers (Fałkowski et al., 2017).

With the deregulation of the dairy market, reduced policy support, and concentrated market power, competitive pressures on milk producers have intensified. This motivates the development of alternative forms of contractual relationships. Collective organisations such as cooperatives can offer

[1] Calculated on AHDB data extracted on the 18 December 2024 from https://ahdb.org.uk/dairy/GB-producer-numbers
[2] Calculated on AHDB data extracted on the 18 December 2024 from https://ahdb.org.uk/dairy/milk-to-feed-price-ratio

farmers improved bargaining power, better access to affordable inputs, and enhanced price stability. Most research on welfare distribution within agricultural supply chains focuses on price transmission (Bakucs et al., 2014), with few studies on contracts' effects on farmer welfare, largely in developing countries (Bellemare and Bloem, 2018). Exceptions include Michalek et al. (2018) on Slovak farms and Verhaegen and Van Huylenbroeck (2001) in Belgium. There are, however, no studies examining collective organisations' impact on the economic performance of UK dairy farmers. Using survey data from 200 dairy farmers in Devon and Somerset and a Coarsened Exact Matching (CEM) approach (Porro and Iacus 2009; Iacus et al. 2012), we estimate the effect of participating in a collective organisation on selling prices, production costs, and markup. Additionally, with penalised regression, we explore which agreement and farmer's characteristics are linked to participating in collective organisations. The findings show that selling through a collective organisation results in 7.3% higher production costs and approximately 23 percentage points lower markups. These increased costs stem from milk collection, transportation, storage, marketing, promotion services, and the production of milk that meets the collective organisation's strict standards. However, these costs are offset by guaranteed market access, stronger bargaining power with retailers and processors, and better alignment with market demand, helping to prevent oversupply and the price volatility that can come with it.

The paper is organized as follows: Section 2 provides a conceptual framework for our research question. Section 3 describes the CEM and the penalised regression methodologies and provides an overview of the data. Section 4 shows the results. Finally, Section 5 provides a discussion of the results and concludes.

2. Conceptual framework

We adopt a transaction cost framework (Williamson 1989) to analyse how farmers' decisions to either participate in a cooperative or sell through the spot market affect their performance. According to this framework, farmers are assumed to choose the option that minimizes their overall transaction costs. Participation in a collective organisation is widely associated with enhanced bargaining power, improved market access, and lower transaction costs – such as those related to searching for buyers, negotiating terms, and enforcing agreements (Alho 2015). In addition, collective arrangements can help reduce market uncertainties. However, while transaction costs may be lower within cooperatives, other types of costs can increase. These may include compliance with higher quality standards, coordination of logistics, and administrative overhead. Despite these additional costs, cooperatives often benefit from greater bargaining power through the pooling of commodities and resources, which can lead to higher farm-gate prices.

Based on this framework, we expect that joining a collective organisation have a positive effect on the price received by farmers. This hypothesis is frequently supported in the empirical literature, although the evidence comes primarily from developing countries (Grashuis and Su 2019). In contrast, the effect on overall costs is less clear, as the reduction in transaction costs may be offset by increased production and coordination expenses. As a result, the net impact on farmers' markup depends on whether the price premium is sufficient to compensate for potential additional costs.

3. Methodology and Data

To reduces the imbalance in covariates between the treated and control groups and to evaluate the differences in outcomes between them, we use the Coarsened Exact Matching (CEM), a counterfactual matching methodology. To match a treated unit with similar control units, the CEM first coarsens each variable into substantively meaningful groups and then performs an exact match based on these coarsened data. Finally, it retains only the original (uncoarsened) values of the matched data. It is relevant to mention that higher coarsening (higher number of strata) is associated with a lower level of imbalance, which comes however at the expense of a higher drop in observations. After preprocessing data with CEM, the analysis can follow using a simple difference in means provided by a linear regression of outcome on treatment (for details see Porro and Iacus 2009; Blackwell et al. 2009, Iacus et al. 2012). A limitation of our data is that we observe covariates only for the survey year, and not before the treatment (i.e. selling through a collective organisation). Therefore, we cannot ensure covariate balance in the pretreatment period, nor we can confirm their independence from participating in a collective organisation. Consequently, this analysis should not be interpreted as causal, but rather as a more accurate estimate compared to a simple comparison between treated and control groups.

In the second step, we investigate agreements and farmers' characteristics associated with participation in a collective organisation using a penalised regression approach. In particular, we apply (logit) LASSO, Elastic Net, and Ridge. Penalised regressions add a 'penalty' to the estimated coefficients, forcing them towards zero (LASSO and Elastic Net) or exactly to zero (Ridge), therefore reducing model complexity and allowing for the selection of explanatory variables. We use these methods to select supply chain features and farmers' characteristics associated with collective organisation agreements among 30 different explanatory variables.

3.1. Data description

Because of the lack of data on SCAs, we use survey data from a small but representative sample of 200 dairy farmers in Devon and Somerset (South West England). The sample includes 88 farms from Somerset and 112 from Devon. The total farm population in the two regions is approximately 1,300. With this sample size, the survey maintains a 5% margin of error (within the commonly accepted threshold ≤10%). In these two counties, there are approximately 20% of all English dairy farmers. The survey was conducted by telephone interviews in November-December 2017 and the covered questions refer to the 2016-17 financial year.

The binary treatment variable distinguishes between farms selling milk exclusively through a collective organisation – cooperatives, producer organisations (PO), inter-branch organisations (IBOs), or farmers' unions and associations – or through individual businesses with a bilateral contract (e.g., retailers). The outcome variables are milk selling price (€/litre), production costs (€/litre), and markup (%). The markup is calculated as follows:

$$markup = \frac{selling\ price\ -\ production\ costs}{production\ costs} \tag{1}$$

To enhance comparability between farms using different sales channels, we consider only those with exclusive contracts (i.e., towards a single collective organisation or individual business). Among the 200 farmers, 173 have an exclusive selling agreement. We considered 4 observations as outliers, i.e. those with a markup lower than the 1^{st} and greater than 99^{th} percentile of the sample distribution. After removing farmers with incomplete data and outliers, the final sample size for the analysis is 169 observations. Summary statistics of all the variables are presented in Table A1 in the Appendix. 43.8% of farms in our sample participate in a collective sale agreement. The average selling price in our sample is about 0.25 euro per litre of milk, while the production cost is about 0.21 €/litre. The average markup is about 36%.

We use the CEM approach to reduce covariates imbalances among farmers. Covariates used for matching are the age (4 categories) and the education of leading farmers (4 categories), total milk production (litres), hectares used for milk production (ha), herd size (number of cows), yield (litre/cow), and full-time workers. In the Appendix, we provide a description of the strategy used to select the "preferred" cutpoints combination used for performing the CEM. Table A2 in the Appendix shows the covariates imbalances before and after matching. We focus on the 'Multivariate L1 distance' that represents the overall imbalance with respect to the full joint distribution (including interactions) of the covariates between the treated and control groups. Panel A shows a value of L1 before matching equal to 0.865. It represents a baseline reference for the unmatched data. A good matching solution would produce a reduction in such a statistic. After considering different alternative cutpoints to define the coarsening, we select those that reduce the the value of the multivariate L1 distance, preserve statistical power, and minimize the loss of observations (see Appendix). Panel B shows a value of L1 after matching equal to 0.637, which is about 23 percentage points lower than the unmatched case (with 107 farmers being matched). Furthermore, Table A3 presents a more standard covariates (average) imbalance and a t-test for the difference in average values that never rejects the hypothesis that they are equal to zero.

4. Results

Table 1 presents our main findings. The estimated coefficient for the treatment variable in column 1 shows that selling milk through a collective organisation rather than to individual businesses is associated with a slightly higher average selling price (0.003 €/litre), although not statistically significant. As shown in column 2, collective sales have statistically significant higher production costs of 0.015 €/litre (about +7.3% of the average costs). This increase in costs leads to a significant reduction in markup of 23 percentage points (column 3).[3] A back-of-the-envelope estimate suggests that the sample average farmer selling exclusively through a collective organisation would face a revenue increase of €5,550 and an increase in total costs of €27,750, based on average annual milk production of 1,850,000 litres. Robustness checks, such as unmatched OLS, the addition of covariates, and a Propensity Score Matching support these findings (see Table A4). Qualitatively similar results are reached using 150 different combinations of cutpoints in the CEM algorithm (see Figure A1).

To understand why SCAs with collective organisations imply higher production costs and lower markup, we further investigate five categories of agreements' characteristics associated with collective organisations using penalised (logit) regression, namely: i) special contract clauses (contract duration, services, producer protections); ii) pricing methods; iii) timing of milk delivery and pricing; iv) costs associated to the agreement; and v) required quality and production standards. Given a large number of variables (30 in total), we apply penalised regression methods (namely, Ridge, LASSO, and Elastic Net) to select those variables with higher predictive power.

Table 1. Regression results after CEM of selling to a collective organisation on farmer's outcomes

	(1) Selling price €/litre	(2) Production costs €/litre	(3) Markup %

[3] The sample size decreases from 169 to 107 observations after applying CEM. This reduction may raise concerns about generalizability and statistical power (see Appendix for further discussion). The statistical (post) power tests for each outcome (price, costs, and markup) are 0.21, 0.72, and 0.82, respectively.

Collective (binary)	0.003	0.015*	-0.233*
	(0.006)	(0.008)	(0.126)
Observations	107	107	107
R^2	0.002	0.030	0.038

Note: the treatment variable 'collective' is = 1 if the exclusive sale is through a collective organisation, 0 otherwise. Standard errors in parentheses are robust to heteroskedasticity. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A4 in the Appendix shows the standardised coefficients of the selected variables associated with the LASSO and Elastic Net estimations, while Figure A1 shows the absolute value of the standardised coefficients from the Ridge estimation. We then run a standard logit model regressing the binary variable that identifies the agreement with a collective organisation on the selected variables. Table 2 shows the estimated coefficients.[4]

Dairy farmers selling through a collective organisation are more likely to get paid after the milk delivery and to receive a price that is based on the yearly organisation's profit. This SCA is also more likely to be associated with logistics (collection, storage, transport, handling, etc.), and marketing costs. Looking at the overall sample (Table 2, column 2), dairy farmers selling through a collective organisation are more likely to comply with higher quality standards for the milk (e.g., taste and nutritional content), increasing production costs. Conversely, they are less likely to receive a price based on production costs, therefore, their milk prices do not follow the volatility of input costs. Finally, they are less likely to comply with animal welfare standards.

Table 2. Logit model using selected variables from LASSO and Elastic Net describing agreements' characteristics associated with collective organisations.

	(1) Matched obs. Collective	(2) Whole sample Collective
Payment after delivery	2.365***	3.195***
	(0.826)	(0.783)
Price varies with the share of profits	1.556***	1.855***
	(0.553)	(0.481)
Logistic costs	1.059**	0.918**
	(0.510)	(0.457)
Marketing costs		0.537

[4] Note that we are not interested in the magnitude of the coefficients or their causal interpretation. We aim to highlight the presence or absence of relevant features in the case of collective organizations.

		(0.445)
Price varies with production costs		-0.957*
		(0.540)
Quality standards		2.057**
		(0.883)
Animal welfare standards		-2.393***
		(0.853)
Observations	97	161
Pseudo R2	0.25	0.34

Note: the dependent variable 'collective' is = 1 if the exclusive sale is through a collective organisation, 0 otherwise. The explanatory variables are all binary variables ('yes' = 1, 'no' = 0). The exact questions are provided in the Appendix. Eight farmers did not answer all the questions included, therefore the number of observations is lower than in Table 1. Standard errors in parentheses are robust to heteroskedasticity. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

To better understand the reasons why dairy farmers sell through collective organisations, we perform an additional analysis on two groups of variables: i) the farmers' perception towards the SCA; ii) the farmers' risk-taking attitude. These variables are measured on a 1-5 Likert scale (1 = strongly disagree, 5 = strongly agree) and they are analysed using a logit regression. Complete questions are available in the Appendix.

Table 3 summarizes the results. Farmers who do not have alternative options are not associated with collective organizations. This implies that farmers do have other opportunities to sell their milk. Similarly, farmers stating that the sale agreement provides more stable prices from year to year than alternative buyers are less likely to rely on collective organisations. This means that cooperatives/PO offer fewer stable prices. However, they offer more opportunities for negotiating prices, an advantage less likely when engaging with individual businesses (e.g., retailers). A negative perception is that farmers who see delays in payments are more likely to be associated with a collective organisation. Looking at the risk attitudes, farmers with higher attitudes towards production risks (such as adopting new milking regimes, new fodder, and new breed of cow), marketing, and pricing risks are less likely to be associated with collective organisations. Contrary, attitudes toward financial risks (e.g. new investments) increase the propensity to sell through a collective organisation. These results suggest that farmers seek stable market access and protection and at this end, they are willing to accept price volatility, higher costs and financial exposure which can be faced through a higher predisposition to taking financial risks.

5. Discussion and conclusion

Sales agreements are fundamental to the economic viability of farms. Using a sample of farmers in Somerset and Devon, England, in this note we present the results of an empirical analysis showing that dairy farmers face a trade-off between benefits and costs when selling through collective organisations. The benefits consist especially in market coordination, through higher negotiating power with respect to retailers and processors, and alignment with market demand to avoid oversupply that can lead to milk price volatility. Without the milk quota system, the pricing mechanism of the collective organisation is the main tool to regulate supply. Different types of collective organisations adopt different pricing mechanisms. A&B pricing mechanisms – farmers receive A prices on defined volumes of milk agreed in advance, while any excess of milk delivered is paid a B price discouraging overproduction – are among the most common tools to keep supply and demand tightly controlled. Larger cooperatives, such as ARLA, buy all the milk delivered but rely on their ability to export to tighten the supply reducing the need for price adjustments and allowing farmers to expand.

Table 3. Farmers' perception and risk attitudes and their association with collective organisations.

	(1) Matched Collective	(2) Whole sample Collective
No alternative options for selling milk	-0.213	-0.156
	(0.168)	(0.122)
Higher prices than alternative buyers	-0.119	-0.013
	(0.231)	(0.165)
More stable prices than alternative buyers	-0.400*	-0.497***
	(0.242)	(0.165)
More possibilities for negotiating prices	0.302	0.292*
	(0.205)	(0.154)
There are delays in the payments	0.166	0.295
	(0.269)	(0.224)
The costs of this sale agreement are too high	0.278	0.087
	(0.237)	(0.184)
Willingness to take production risks	-0.266	-0.382**
	(0.226)	(0.173)
Willingness to take marketing and price risks	-0.191	-0.062
	(0.248)	(0.187)
Willingness to take financial risks	0.600***	0.487***
	(0.230)	(0.168)
Observations	91	154

Pseudo R^2	0.12	0.12

Note: Explanatory variables are on a 1-5 Likert scale. The exact questions are provided in the Appendix. Ten farmers did not answer all the included questions, therefore the number of observations is lower than in Table 1. Standard errors in parentheses are robust to heteroskedasticity. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

However, these benefits come with costs that with deregulation have been partially shifted from public support to farms. Farmers pay for services such as milk collection, transport and storage, and for marketing and promotion. Moreover, farmers make products to the collective organisation's specifications, owing full responsibility for the competencies and capital needed for the process technology. Indeed, farmers selling through collective organisations show a higher propensity to take financial risks, such as taking loans for investments. However, these costs are the price farmers pay for having some confidence that their milk supply has a buyer and a decent price, instead of having to hope for a spot contract most probably not aligned with retailers' price.

The shift of costs from public support to individual farms highlights the need for nuanced policy that account for the dual role of collective associations as both market coordinators and informal insurance mechanisms. While farmers face higher costs and often receive lower prices due to stricter quality requirements, CAs offer significant value by reducing exposure to price volatility, thus stabilising farm income and lowering financial risk. Policymakers should recognise this implicit insurance function and consider supporting it through targeted instruments, such as subsidies for compliance with quality standards, investment support for required technologies, or risk-sharing mechanisms like income stabilization tools.

While this note provides some evidence of the trade-offs farmers face with respect to selling practices and how SCAs affect their economic performance in a developed country, it has important limitations. First, the estimates cannot be considered causal, as we lack information on pre-treatment periods. Second, the results are based on a limited sample of farms from only two counties in England. Therefore, the findings cannot be generalized to the national level, nor to other developed countries. Finally, we are unable to clearly identify the mechanisms that lead farmers to participate in collective organizations, even when it may result in a lower markup. Further research is needed to explore both collective and contractual arrangements in modern agricultural markets. A critical step in this direction would be to improve the current lack of data on SCAs, with larger samples across multiple sectors and countries to analyse.

References

Alho, E. (2015). Farmers' self-reported value of cooperative membership: evidence from heterogeneous business and organization structures. *Agricultural and Food Economics*, *3*(1). https://doi.org/10.1186/s40100-015-0041-6

Bakucs, Z., Fałkowski, J., Ferto, I. (2014). Does Market Structure Influence Price Transmission in the Agro-food Sector? A Meta-analysis Perspective. *Journal of Agricultural Economics*, 65(1): 1–25. https://doi.org/10.1111/1477-9552.12042

Bellemare, M. F., Bloem, J.R. (2018). Does contract farming improve welfare? A review. *World Development*, 112: 259–271. https://doi.org/10.1016/j.worlddev.2018.08.018

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). cem: Coarsened exact matching in Stata. In *The Stata Journal* (Vol. 9, Issue 4).

Fałkowski, J., C. Ménard, R.J. Sexton, J. Swinnen and S. Vandevelde. (2017), Unfair trading practices in the food supply chain: A literature review on methodologies, impacts and regulatory aspects, European Commission, Joint Research Centre.

Grashuis, J., & Su, Y. (2019). A review of the empirical literature on farmer cooperatives: performance, ownership and governance, finance, and member attitude. *Annals of Public and Cooperative Economics*, *90*(1), 77–102. https://doi.org/10.1111/apce.12205

Iacus, S.M., King, G., Porro, G., (2012). Causal Inference Without Balance Checking: Coarsened Exact Matching. *Political Analisis*, 20(1), 1-24. https://doi.org/10.1093/pan/mpr013

Michalek, J., Ciaian, P., Pokrivcak, J. (2018). The impact of producer organizations on farm performance: The case study of large farms from Slovakia. *Food Policy*, 75: 80–92. https://doi.org/10.1016/j.foodpol.2017.12.009

Porro, G., & Iacus, S. M. (2009). Random Recursive Partitioning: A matching method for the estimation of the average treatment effect. *Journal of Applied Econometrics*, *24*(1), 163–185. https://doi.org/10.1002/jae.1026

Sexton, R.J. (2012). Market power, misconceptions, and modern agricultural markets. *American Journal of Agricultural Economics*, 95(2): 209–219. https://doi.org/10.1093/ajae/aas102

Verhaegen, I., Van Huylenbroeck, G. (2001). Costs and benefits for farmers participating in innovative marketing channels for quality food products. *Journal of Rural Studies*, 17: 443–456. https://doi.org/10.1016/S0743-0167(01)00017-1

Williamson, O. E. (1989). Transaction cost economics. In *Handbook of Industrial Organization* (Vol. 1, pp. 135–182). Elsevier Science Publishers. https://doi.org/10.1016/S1573-448X(89)01006-X

Supplementary Materials

Table A1. Summary statistics for outcome, treatment, and control variables

	N	Mean	SD	Min	Max
Collective sale (dummy)	169	0.438	0.498	0	1
Average selling price (€/litre)	169	0.248	0.032	0.18	0.41
Production costs (€/litre)	169	0.205	0.050	0.053	0.324
Markup (%)	169	0.355	0.731	-0.231	3.348
Age (4 categories)	169	2.71	0.876	1	4
Education (4 categories)	169	2.793	0.723	1	4
Hectares for milk production (ha)	169	141.189	152.642	4	1011
Total milk production (litre)	169	1851650	2683897	20000	23000000
Herd size (number of cows)	169	234.142	214.657	17	1600
Yield (litre/cow)	169	7960.009	1886.203	999	15000
Full-time workers (number of)	169	3.698	3.512	0	32

Note: Authors calculation on survey data. 'Age' is divided into four categories (1-4): less or equal to 40 years, between 41 and 50, between 51 and 65, and more than 65 years. 'Education' is divided into four categories (1-4): primarily, lower secondary, higher secondary, and university. The variable 'hectares for milk production' refers to the total area cultivated for milk production (including land used for growing fodder crops).

Table A2. Covariates imbalances before (A) and after (B) Coarsened Exact Matching

Panel A

Multivariate L1 distance: 0.865

Univariate imbalance:

	L1	mean	min	25%	50%	75%	max
Age	0.1778	0.2033	0	1	0	0	0
Education	0.0514	0.0078	1	0	0	0	0
Full-time workers	0.1932	0.2964	-1	0	0	0	8
Herd size	0.1316	24.749	6	-10	27	5	200
Hectares	0.0176	17.741	9	-7	-10	0	364
Production	0.0815	4.9e+05	55000	-1.2e+05	0	0	8.0e+06
Yield	0.1737	-289.73	-2801	-700	-500	0	3000

Panel B

Multivariate L1 distance: 0.637

Univariate imbalance:

	L1	mean	min	25%	50%	75%	max
Age	0.0648	-0.0648	0	0	0	0	0
Education	0.0833	0.0401	0	0	0	0	0
Full-time workers	0.0432	0.1049	0	0	0	0	8
Herd size	0.0741	4.292	23	-10	15	-5	200
Hectares	0.1512	2.394	9	2	-5	-11	364
Production	0.1142	5.9e+05	-18000	41000	0	-2.0e+05	8.0e+06
Yield	0.0694	91.88	-300	-3	0	0	3000

Note: Panel A shows the multivariate and univariate imbalance of covariates without matching observations. Panel B shows the multivariate and univariate imbalance after implementing the CEM algorithm. The cutpoints for the covariates are: all four categories for 'age'; category 1 for 'education'; 4 workers for the variable 'Full-time workers'; 200 cows for 'herd size'; 141.2 hectares for 'hectares for milk production'; 800,000 litres for 'production'; 6666, 8000, and 9000 litre/cow for 'yield'.

Table A3. Average covariances imbalance before and after CEM and t-test on the difference

	Non-matched			Matched		
	Mean		Difference	Mean		Difference
	Individual	Collective		Individual	Collective	
Age (4 categories)	2.62	2.82	-0.20	2.60	2.63	-0.03
			(0.14) [0.13]			(0.15) [0.87)
Education (4 categories)	2.79	2.80	-0.01	2.77	2.81	-0.04
			(0.11) [0.94]			(0.13) [0.75]
Full-time workers	3.57	3.86	-0.30	3.55	3.83	-0.29
			(0.55) [0.59]			(0.81) [0.72]
Ha for milk production	223	248	-25	226	237	-11
			(33) [0.46]			(48) [0.82]
Herd size (n. of cows)	133	151	-18	123	133	-11
			(24) [0.46]			(30) [0.73]
Milk production (litre)	1637568	2126487	-488918	1748755	2340593	-591838
			(415658) [0.24]			(636183) [0.35]
Yield (litre/cow)	8087	7797	290	8273	8267	5
			(292) [0.32]			(357) [0.99]

Note: 'Individual' and 'Collective' represent farms that sell milk exclusively through individual businesses with bilateral contracts (e.g., retailers) or through collective organizations (e.g., cooperatives). Values in curly brackets indicate standard errors from a two-sample t-test, while values in square brackets represent p-values for the alternative hypothesis that the difference is not equal to zero.

Table A4. Alternative results of selling through collective organizations on farmer's outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	Unmatched OLS	Unmatched OLS	Unmatched OLS	Unmatched OLS	CEM	PSM
Dependent variable:						
Average selling price	0.002	0.00001	0.004	0.001	0.003	0.003
(€/litre)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)
Production costs	0.010	0.009	0.012	0.011	0.015	0.012*
(€/litre)	(0.007)	(0.008)	(0.007)	(0.008)	(0.009)	(0.007)
Markup (%)	-0.236*	-0.251*	-0.183*	-0.204*	-0.229*	-0.186*
	(0.132)	(0.140)	(0.107)	(0.107)	(0.126)	(0.108)
Observations	196	196	169	169	107	169
Exclusive sale only			x	x	x	x
Covariates		x		x	x	

Note: the treatment is the variable 'collective' that is a binary variable equal to 1 if the sale is through a collective organisation, 0 otherwise. 'Exclusive sale only' indicates that farmers sale only through a single agreement. Standard errors (in parenthesis) are adjusted for heteroskedasticity in the case of OLS and CEM, while produced by a bootstrap procedure (100 times) in the case of Propensity Score Matching (PSM) to account for the fact that the propensity score is estimated. Significance levels: * p<0.1, ** p<0.05, *** p<0.01

Table A5. LASSO and Elastic Net standardised coefficients of selected variables via Cross-Validation (10-folds).

		(1)	(2)	(3)	(4)
		LASSO	Elastic Net	LASSO	Elastic Net
		Full sample	Full sample	Matched	Matched
Payment entirely after delivery	(QC26)	0.788	0.759	0.471	0.453
Price varies with the share of profits	(QC21)	0.545	0.531	0.379	0.368
Logistic costs	(QC31)	0.243	0.239	0.239	0.235
Marketing costs	(QC32)	0.098	0.101		
Price varies with production costs	(QC17)	-0.165	-0.161		
Quality standards	(QC36)	0.068	0.064		
Animal welfare standards	(QC39)	-0.030	-0.028		

Note: the dependent (dummy) variable is ‘collective’. The definition of each explanatory (dummy) variable is provided in the text below.

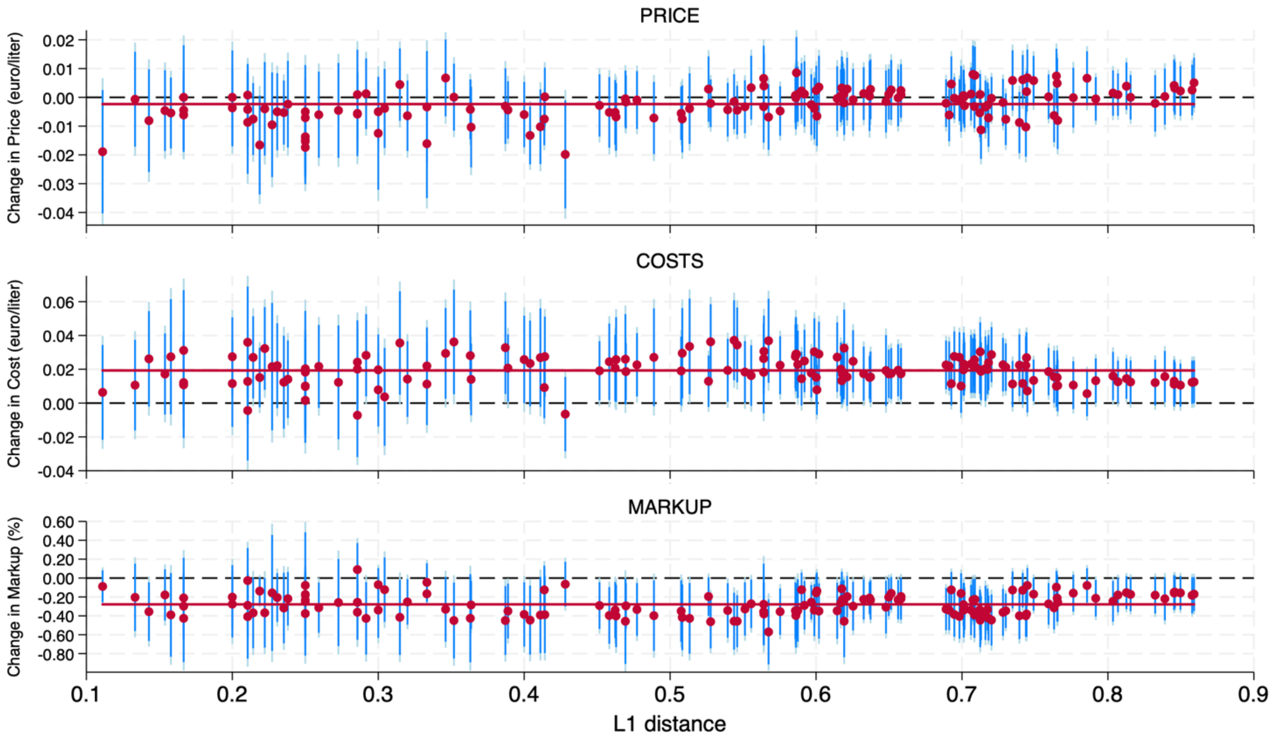

Figure A1. Estimated coefficients after different CEM combinations. The horizontal axis represents the corresponding L1 statistics. Red dots represent point estimates, while blue bars are the 90 and 95 confidence intervals. The red lines are the average of the estimated coefficients (-0.002 for price, 0.019 for costs, and -0.278 for markup).

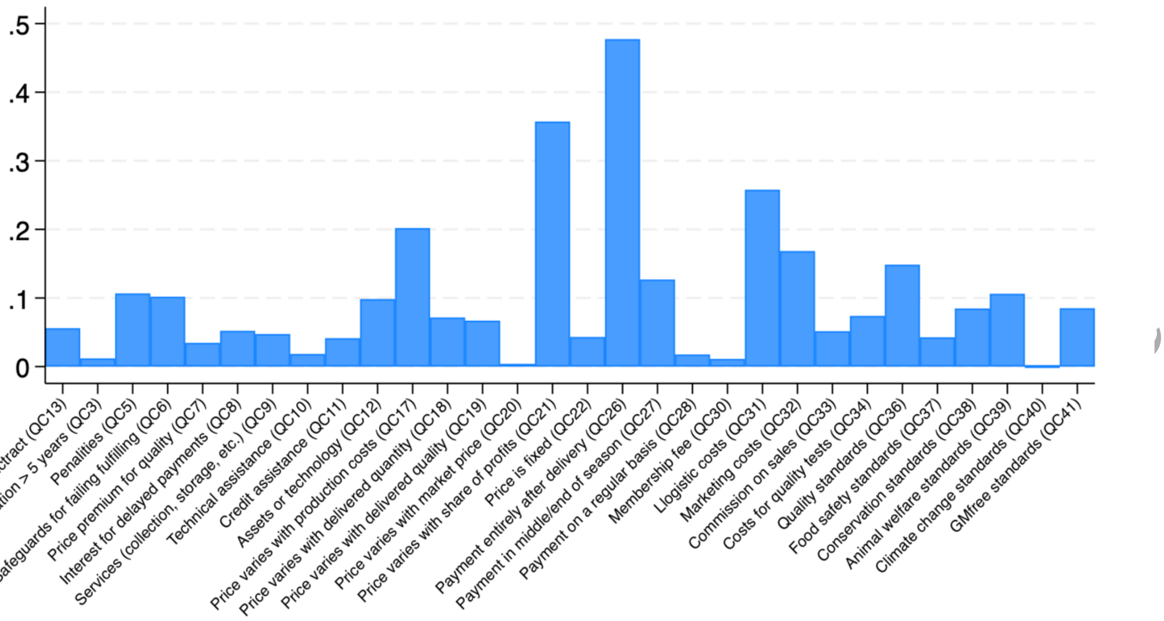

Figure A2. Standardised absolute values of the coefficients from Ridge estimation

Definition of the variables selected by the LASSO and Elastic Net (Table 2 and Table A5)

Variable: "Payment entirely after delivery"

"According to the agreement of this sale/membership rules of a collective organization, when do you get paid? Please select one option:"

QC.26	Entirely after the delivery of products	Yes	No

Variables: "Price varies with the share of profits"; "Price varies with production costs"

"On what basis is the price of milk determined by the agreement?"

QC.21	Variable price based on share of organization's profit	Yes	No
QC.17	Variable price based on production costs	Yes	No

Variables: "Logistic costs"; "Marketing costs"

"In this sale agreement, which of the following costs do you incur?"

QC.31	Collection, storage, transport, handling, etc	Yes	No

QC.32	Promotional and marketing costs	Yes	No

Variables: "Quality standards"; "Animal welfare standards"

"Does the buyer/collective organization require specific production/quality standards that you have to comply with?"

QC.36	Standards on the quality of the final product(s) (taste, colour, shape, nutritional content, chemical composition, etc…)	Yes	No
QC.39	Standards on animal welfare	Yes	No

Definition of the variables presented in Table 3

"On a scale from 1 (strongly disagree) to 5 (strongly agree) how much do you agree with the following statements regarding your satisfaction with respect to this sale agreement?"

QC.44	I do not have any alternative options to sell my products
QC.45	This sale agreement provides higher prices than alternative buyers
QC.46	This sale agreement provides more stable prices from year to year than alternative buyers
QC.47	This sale agreement provides more possibilities for negotiating prices
QC.48	There are delays in the payments
QC.49	The costs associated with this sale agreement are too high (e.g. storage, marketing and promotion, commission on sales)
QX.15	I am willing to take more risks than my colleagues with respect to dairy production (e.g. adopt new milking regimes, new fodder, new breed of cow)
QX.16	I am willing to take more risks than my colleagues with respect to marketing and pricing risks (e.g. experiment with new sales channels, and new markets)
QX.17	I am willing to take more risks than my colleagues with respect to finance (e.g. loans for production facilities, new investments)

Description of the CEM implementation and cutpoints selection

To investigate the impact of coarsening cutpoints in the implementation of Coarsened Exact Matching (CEM), we run the algorithm multiple times using different sets of cutpoints. Specifically, we explore various combinations of cutpoints as follows.

First, consider the histograms in Figure A3. We observe that most variable distributions are right-skewed, and using too many thresholds would result in many empty or sparsely populated strata. To avoid this, we use the mean, median, mode, and quartiles of each variable's distribution as candidate cutpoints. For categorical variables such as ‘age’ and ‘education’, which take discrete values from 1 to 4, we use the actual values instead of quartiles.

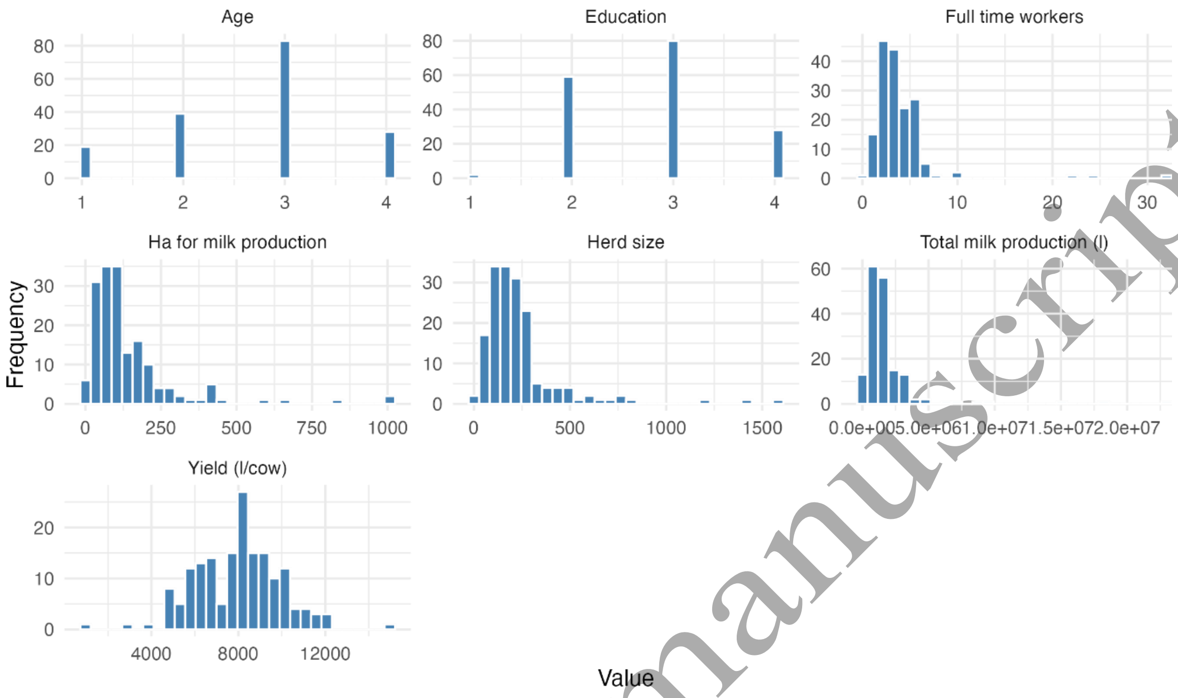

FIGURE A3. Histograms of the variables used in the CEM.

In total, we evaluate all 75000 unique combinations of these cutpoints and report the corresponding L1 statistics in Figure A4, panel A. For each group of matched observations (e.g., those resulting in 50 matched units), we retain only the combination that yields the lowest L1 statistic. We discard combinations with L1 values greater than 0.86, which corresponds to the L1 value in the unmatched case. This filtering process reduces the number of relevant combinations from 75000 to 151, shown in panel B.

As illustrated in both panels of Figure A4, there is a clear positive relationship between the L1 statistic and the number of matched observations: lower L1 values tend to correspond to smaller matched samples, and vice versa. This tradeoff has implications for the statistical power and internal validity of our estimates. A smaller number of matched observations typically implies lower statistical power—that is, a reduced probability of rejecting the null hypothesis (β = 0) when the alternative hypothesis (β ≠ 0) is true. Moreover, matching on a small subset of the original sample may compromise the internal validity of the estimates, as they are based on a less representative subset.

These considerations suggest that relying solely on the CEM combination that minimizes the L1 statistic is not optimal in this context. Instead, we select our preferred combination by jointly considering the L1 statistic, the number of matched observations, and the statistical power of the resulting estimate.

A. B.

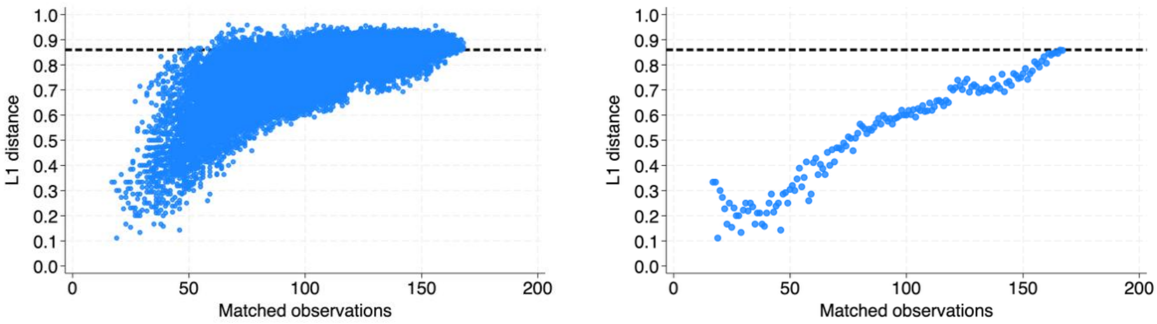

FIGURE A4. Scatter plots showing the relationship between the L1 distance statistic and the number of matched observations after applying CEM. Panel A displays all 75000 combinations of cutpoints, while Panel B shows the subset of combinations that minimize the L1 statistic for each fixed number of matched observations. The dashed line represents the reference L1 value (0.86) obtained without applying any matching.

Figure A5 illustrates this selection process. Panel A displays the L1 statistic, the estimated coefficient (with markup as the dependent variable), and the number of matched observations (represented by bubble size). All but one of the 151 combinations yield negative coefficients; however, those with the lowest L1 values typically correspond to small sample sizes. We select the combination highlighted with a red circle, which substantially reduces the L1 statistic (from 0.865 to 0.637) without excessively reducing the number of matched observations. Notably, the estimated coefficient from this combination is close to that obtained from the unmatched OLS (red cross), suggesting consistency with the full-sample estimate.

In panel B, we demonstrate that this selected estimate has sufficient statistical power (above 80%, indicated by the dashed line). It also outperforms alternatives with fewer observations or coefficients that diverge significantly from the full-sample estimate. Importantly, choosing similar combinations yields comparable results (see Figure A1), indicating that our findings are robust to reasonable variations in the choice of cutpoints.

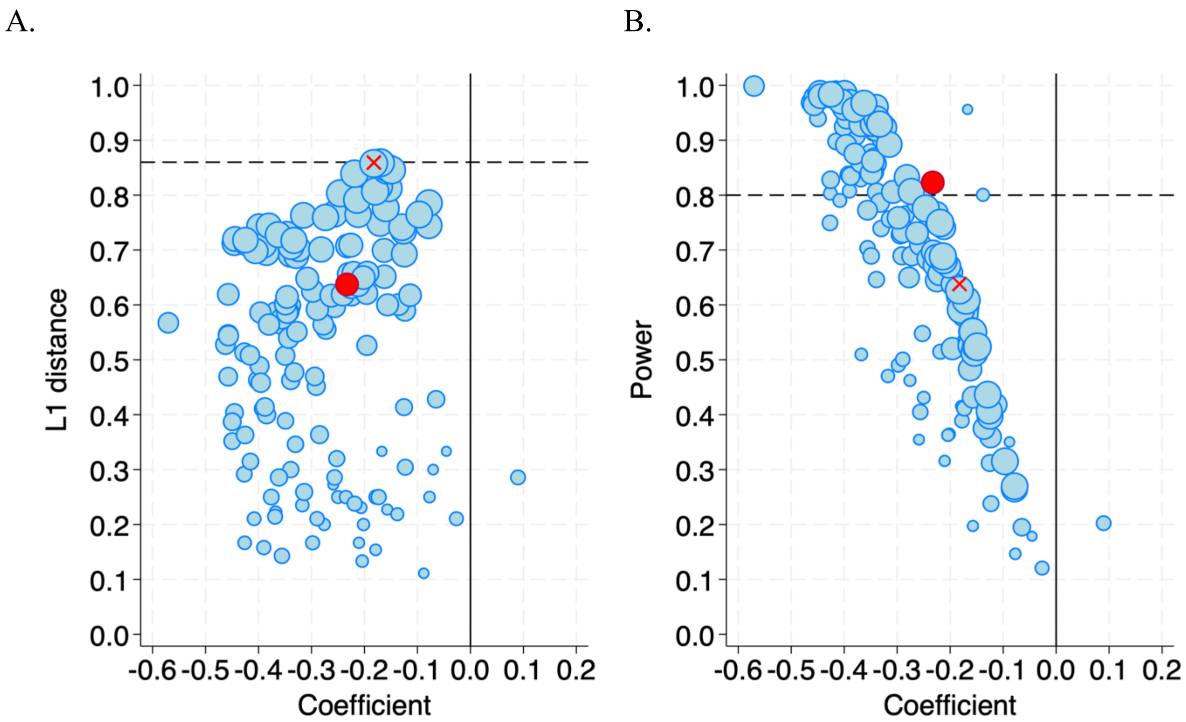

FIGURE A5. Scatter plots of the estimated markup coefficient against the L1 distance statistic (Panel A) and the statistical power (Panel B). Bubble size is proportional to the number of matched observations. The red circle marks the preferred CEM combination, while the red cross represents the unmatched OLS estimate. In Panel A, the dashed line indicates the reference L1 value (0.865) obtained without applying cutpoints. In Panel B, the dashed line marks the commonly used threshold for sufficient statistical power (80%). The L1 statistics are obtained directly from the *cem* command in Stata, while power statistics are calculated using the *power oneslope* command after each regression.