ARTICLE

The role of animal breeding in productivity growth: Evidence from Wisconsin dairy farms

Jared Hutchins[1] | **Yating Gong**[2] | **Xiaodong Du**[2]

[1]Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA

[2]Department of Agricultural and Applied Economics, University of Wisconsin-Madison, Madison, Wisconsin, USA

Correspondence
Jared Hutchins, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaig, Champaign, Illinois, USA.
Email: jhtchns2@illinois.edu

Abstract

We examine the relationship between investments in animal breeding and productivity growth on Wisconsin dairy farms using a control function approach. We incorporate farm-level annual investment in breeding and genetics into the law of motion of productivity as in De Loecker (2013) to test the relationship between these investments and realized productivity. Our unique dataset also allows us to look at the effect of choosing bulls with high milk yield potential on productivity. Our results indicate that breeding investments made 3 years prior are associated with higher productivity of the current cohort. However, the farms with the highest level of productivity reap the lowest benefits from breeding investments, suggesting that there are diminishing returns to investing in genetics. When milk output is not quality adjusted, the contribution of breeding to productivity is undetectable, suggesting that breeding and investments in milk quality are related. We conclude that investments in breeding and genetics significantly contribute to dairy farm productivity, especially in terms of milk quality.

KEYWORDS
control function, genetics, milk quality, production function

JEL CLASSIFICATION
C25, C26

1 | INTRODUCTION

Milk yield has experienced significant growth in the past decades. Figure 1 charts the growth path of average cow milk yield from 2003 to 2019, which grew on average 1.4% a year during this period. At
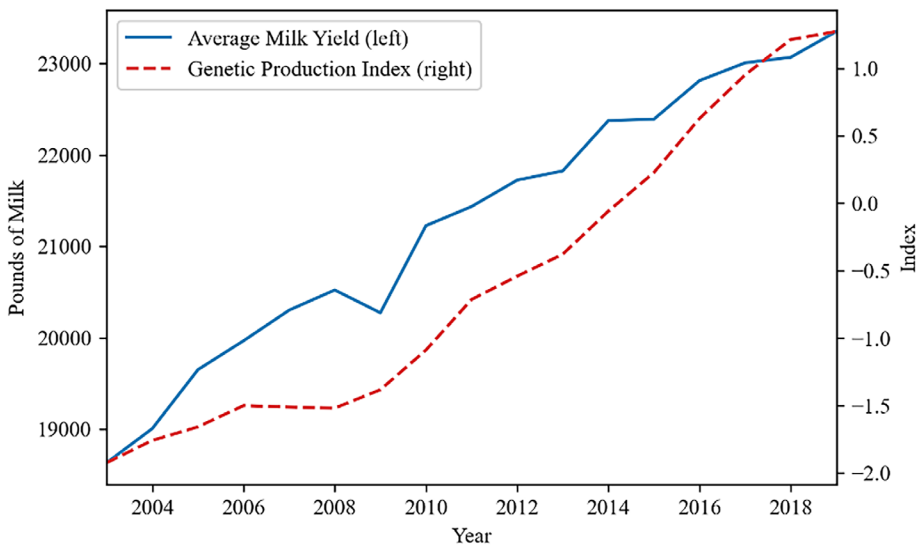
FIGURE 1 Average milk yield and genetic productivity, 2003–2019

the same time, the productivity of dairy bulls available on the market has grown at a higher rate: milk yield, butterfat yield, and protein yield of dairy bulls grew on average 3.3%, 5.1%, and 4% per year over the same period. When combined into a genetic production index,[1] its growth path is similar to that of milk yield. The dairy science literature attributes as much as 50% of the growth in milk yield in the past decades to genetic improvement in dairy cows (Pryce & Veerkamp, 2001; Shook, 2006; Van Raden, 2004). Although its precise contribution is difficult to quantify, it is clear that genetic improvement is an essential vector of productivity growth in dairy.

Choosing genetics is critical for productivity improvement of a dairy herd, and yet investments in herd genetics has been largely ignored when estimating dairy farm production functions (e.g., Jang & Du, 2019; Mukherjee et al., 2013; Njuki et al., 2020). By omitting this decision, conventional production function estimation may misattribute this productivity growth to other input factors, which results in biased input coefficients (De Loecker, 2013). Moreover, not understanding this vital vector of productivity growth in dairy farming makes it difficult to understand the effect of new genetic improvements on the future of the dairy industry.

This paper investigates the effect of genetics investment on dairy farm productivity in Wisconsin. Dairy farms choose bull genetics using artificial insemination (AI) each year in order to change the future productivity of the herd. These investments are reflected in both how much farms spend on breeding each year and the genetic indices of the bulls they choose in each year. In this paper, we leverage both of these measures to better understand how investments in genetic improvement impact the evolution of productivity on dairy farms.

Using two rich data sources on observed farm- and animal-level decisions, we integrate investment in genetic improvements into the dynamics of productivity, which is modeled by a first-order Markov process in the style of Levinsohn and Petrin (2003) and De Loecker (2013). Since investments in genetics impact productivity 3 years from the date of investment, variation in breeding investment helps identify the parameters in the farm production function. Using this method, we analyze both the relationship of genetic investment to productivity growth as well as the extent to which omitting investment in genetics biases estimated factor shares in the production function.

---

[1]This index is a weighted average of milk production, fat production, and protein production traits of dairy bulls for sale. The precise weights are determined by principal component analysis, the details of which are found in the supplementary Appendix A.

We find that both breeding expenditures and the production potential of the genetics farmers choose are significant factors in the dynamic evolution of productivity. In our sample of Wisconsin dairy farms, farmers with higher breeding expenditures 3 years prior had higher productivity, as did farms that chose bulls with a higher genetic production index. We also find that dairy farms with high productivity reap the smallest gains from increasing investment in high-yield genetics, suggesting diminishing returns to investment in genetics. The magnitude of the relationship between breeding investments and productivity depends on whether the output we use is adjusted for milk quality, which suggests that breeding investments and milk quality improvements are related. In addition to providing evidence on how genetics impact productivity, our analysis demonstrates the importance of adjusting for output quality when analyzing dairy farm performance.

Our work expands on the productivity literature concerning dairy by modeling breeding as an endogenous investment into future productivity. The effect of breeding on productivity growth has thus far only been explored as an exogenous factor in technological progress and not as an investment decision. Townsend and Thirtle (2001) calculate that animal breeding research has a 35% return on investment in South Africa using time series data on R&D expenditures. Studies examining the returns to breeding with farm-level data are even more rare. Like the majority of the dairy economics literature, these studies use the distance function approach (Bravo-Ureta et al., 2021). Using a modified Malmquist productivity index, Atsbeha et al. (2012) find that 19% of the productivity growth rate in the Icelandic dairy sector is due to changes in the average sire "breeding value," or estimated productivity. Importantly, Atsbeha et al. (2012) view genetics not as an input that farmers invest in but rather feature of the technology, similar to technological progress. Roibas and Alvarez (2010) use a similar, frontier-based method and find that the most technically efficient dairy farms experience the highest returns to genetics.

The weakness of these approaches is that it does not control for the endogeneity of the genetics of the herd. The genetic productivity of a herd is not an exogenous factor but rather a result of investment decisions made by the farm in the past. The decision to adopt improved genetics is an investment into future productivity based on current conditions including productivity of the existing cohort. Much like the framework of Olley and Pakes (1996), dairy farms make investments every year in the genetics of their herd. Every year, farmers choose bulls to breed with their cows to obtain replacement cows roughly 3 years in the future. These investments in bull quality provide the proxy variable we can use to identify production function coefficients (e.g., Levinsohn & Petrin, 2003; Olley & Pakes, 1996). Our work joins other studies that use the control function approach to correct for the endogeneity of input decisions in dairy farming (Frick & Sauer, 2018; Jang & Du, 2019; Kirwan et al., 2012; Läpple et al., 2021). Using the control function approach, our work represents animal breeding more accurately as an endogenous investment made by the farmer to improve future productivity and not simply as a exogenous feature of the technology. Our work also takes advantage of cow-level data that provide more detailed and granular data for understanding the economics of dairy operations, especially the selection of genetics (Hutchins et al., 2021; Hutchins & Hueth, 2021).

Our work builds on previous control function analyses of dairy by modeling the relationship between genetics investment and productivity by including genetic investments in the law of motion of productivity. De Loecker (2013) demonstrates that, in the case of exporting firms, neglecting critical variables in the law of motion leads to poorly estimated factor shares. In the case of dairy, by omitting intertemporal breeding investments we may be misattributing output to various factors such as labor, feed, or even capital. Our study is also the first to examine the impact of genetic investments in health on productivity, a potentially important but so far unexplored aspect of productivity growth (Townsend & Thirtle, 2001). Finally, our paper contributes to the production function literature by demonstrating an intuitive use for supervised machine learning in model estimation. We correct for survival bias using the method outlined in Olley and Pakes (1996) but significantly improve the probability prediction step by using a random forest algorithm.

The paper proceeds as follows. First, we explain our empirical model that draws on De Loecker (2013) to incorporate genetic investments into the production function. We then describe our three data sources that provide farm- and cow-level information to estimate our model. We present the results of our analysis in the fourth section before concluding with a discussion of the significance of the results.

2 | EMPIRICAL MODEL

In our model, farm i produces Q_{it} units of quality-adjusted milk in period t using these inputs: cows (denoted by C hereafter), labor (L), capital (K), feed (F), and intermediate inputs (M). The output is adjusted with the percentage composition of fat, the percentage composition of protein, and the average somatic cell count (SCC).[2] Labor corresponds to hired workers; capital refers to building, machinery, and equipment; and intermediate inputs refer to electricity. Feed includes total amount of purchased feed.[3] The Cobb–Douglas production function is specified as[4]:

$$Q_{it} = C_{it}^{\beta_c} L_{it}^{\beta_l} K_{it}^{\beta_k} F_{it}^{\beta_f} M_{it}^{\beta_m} exp(\omega_{it} + \varepsilon_{it}), \qquad (1)$$

where ω refers to the Hicks-neutral farm-level productivity measure, which implies that technical change affects milk yield only by changing ω and not the balance of other inputs.[5] The term ε is an i.i.d. measurement error term that accounts for random productivity shocks. The logarithmic form of Equation (1) is:

$$q_{it} = \beta_c c_{it} + \beta_l l_{it} + \beta_k k_{it} + \beta_f f_{it} + \beta_m m_{it} + \omega_{it} + \varepsilon_{it}, \qquad (2)$$

where lowercase letters denote the log form of the uppercase letters, that is, $q_{it} \equiv log(Q_{it})$, $c_{it} \equiv log(C_{it})$, $l_{it} \equiv log(L_{it})$, $k_{it} \equiv log(K_{it})$, $f_{it} \equiv log(F_{it})$, and $m_{it} \equiv log(M_{it})$. To correctly estimate the parameters of this equation, we must address three things: input endogeneity, the quality of output, and survival bias.

2.1 | Input Endogeneity

Input endogeneity is a result of the correlation between farmers' input decisions and unobserved productivity ω_{it}. Therefore, OLS estimates of the input coefficients are biased. There are some approaches proposed in the early literature to deal with endogeneity such as the fixed effect model (e.g., Hoch, 1955), utilizing the first order conditions of flexible inputs (e.g., Hall, 1988), and using input prices as instrumental variables (e.g., Griliches & Mairesse, 1998). The fixed effect model assumes a time-invariant productivity shock, that is, $\omega_{it} = \omega_i$, which does not work well in practice. Similarly, the static first-order condition typically does not hold in the choices of flexible inputs. Using input prices as IVs is also questionable for reasons like not capturing input quality and/or lack of variations across firms.[6] The caveats of the prior approaches motivate the Olley-Pakes/Levinsohn-

---

[2]We discuss the detailed procedure of quality adjustment in a later section.

[3]The construction details of the output and input variables are described in the data section.

[4]We follow the Cobb–Douglas specification of the production function mainly because it is a predominant specification in the industrial organization literature (De Locker & Syverson, 2021). Other specifications include, for example, translog (e.g., De Loecker & Warzynski, 2012) and the constant elasticity of substitution (e.g., Grieco et al., 2016).

[5]Hicks-neutral productivity is defined as a non-input-related or factor-neutral shift of the production function. It measures the changes in output that are not explained by changes in inputs. Therefore, a higher productivity here means more output can be produced from a fixed set of inputs. This is a predominant setup in the industrial organization literature (De Locker & Syverson, 2021).

[6]We refer readers to Section 2.1, Ackerberg et al. (2015) for a detailed discussion of the related methods in the early literature.

Petrin (Levinsohn & Petrin, 2003; Olley & Pakes, 1996) method that we adapt to deal with the endogeneity problem.

OP/LP methods use a proxy variable that is monotonically increasing in productivity, such as investment (in OP) and intermediate inputs (in LP), to control for unobserved productivity. In our study, we use the investments in breeding and genetics, g_{it}, as the proxy for unobserved productivity: $g_{it} = g_t(\omega_{it}, c_{it}, k_{it})$.[7] If there is a positive, monotonic relationship between the amount of investment g_{it} and productivity ω_{it} for any combination of c_{it} and k_{it}, meaning ω_{it} is increasing in g_{it}, then the inverse, $\omega_{it} = g_t^{-1} = \phi_t(g_{it}, c_{it}, k_{it})$, can be used as a valid proxy for productivity. We use two variables to represent for genetic investment: breeding expenditures and the average productivity and health indices of the bulls chosen for breeding. By using g_{it} as the proxy for unobserved productivity, we replace ω_{it} in Equation (2) by $\phi_t(g_{it}, c_{it}, k_{it}) = h_t(g_{it}, c_{it}, k_{it}) + \beta_k k_{it} + \beta_c c_{it}$ and rewrite the production function as follows:

$$q_{it} = \beta_l l_{it} + \beta_f f_{it} + \beta_m m_{it} + \phi_t(g_{it}, c_{it}, k_{it}) + \varepsilon_{it}, \tag{3}$$

where ϕ_t is a third-order polynomial approximation of productivity, which includes the linear terms $\beta_k k_{it}$ and $\beta_c c_{it}$. Based on Equation (3), we obtain the predicted output by regressing observed output on all of the inputs and the proxy variable of breeding investment. Using the predicted output $\widehat{q}$, we estimate the productivity shock $\omega_{it} = q_{it} - \widehat{q}_{it}$ to estimate the parameters of the law of motion of productivity.

We depart from the standard OP/LP model by following De Loecker (2013) and amending the law of motion for productivity. Instead of a first-order Markov process for productivity defined in the standard OP/LP model, we assume that genetic investments in year $t-3$ impact productivity in year t. This time lag is a result of the biological constraints of dairy farming. When a cow becomes pregnant after breeding, it gives birth 10 months later. That offspring is then bred at about 1 year old so that it can begin producing at 2 years old. Altogether, the breeding decision and the productivity resulting from that decision are about 3 years apart. If ξ_{it} denotes the productivity shock, the corresponding law of motion becomes:

$$\omega_{it} = \varphi_t(\omega_{i,t-1}, G_{i,t-3}) + \xi_{it}. \tag{4}$$

For Equation (4), we consider two cases. In the first case, productivity follows an AR(1) process and is a linear function of genetic choice 3 years ago and a productivity shock. It is represented as:

$$\omega_{it} = \rho \omega_{i,t-1} + \gamma G_{i,t-3} + \xi_{it}. \tag{5}$$

In the second case, we add an interaction between the level of productivity last period ($\omega_{i,t-1}$) and the level of investment that will be realized next period ($G_{i,t-3}$)[8]:

$$\omega_{it} = \rho \omega_{i,t-1} + \theta_1 G_{i,t-3} + \theta_2 G_{i,t-3} \omega_{i,t-1} + \xi_{it}. \tag{6}$$

In De Loecker (2013), the interactions allow firms with different levels of productivity to have different returns to exporting. In our case, it captures heterogeneous returns to breeding investments by allowing returns to vary with farm productivity. In essence, $\omega_{i,t-1}$ reflects the productivity of last year's cow cohort and ρ reflects how much those effects persist. The genetic investment $G_{i,t-3}$ reflects the productivity of the incoming cohort, with θ_1 measuring the incoming cohort's impact on

---

[7]Note that the input factors, number of cows (c) and capital (k) are fixed and thus affected only by information at $t-1$. Also, cow number is relevant in the sense that genetic investment g is the total investment on the herd.

[8]This requires the assumption that the new cohort is genetically related to the current cohort. This is reasonable as a majority of Wisconsin dairy farmers raise replacement heifers by themselves for obtain replacements cheaper than in the market and have more control over genetics and disease. We thank a reviewer for pointing this out.

productivity. The effect of adding the new cohort to the existing cohort is measured by θ_2. Because the new cohort is almost always genetically related to the current cohort, the interaction between current and incoming cohorts is a significant part of how productivity evolves on dairy farms. The sign of θ_2 also sheds light on whether there are decreasing or increasing returns to genetic investments in dairy. A positive sign would indicate that the most productive dairy farms get the most out of genetics investment (increasing returns), whereas a negative sign would indicate that the returns to genetics are lowest for the most productive farms (decreasing returns).

Following De Loecker (2013), we assume $E[G_{i,t-3}\xi_{it}] = 0$ to obtain identification. This assumes that breeding decisions made 3 years prior are independent of the evolution of current productivity. This assumption gives a moment condition we can use to identify the parameters of interest by applying the general method of moments (GMM):

$$E\left\{\xi_{it}\left(l_{i,t-1}, k_{i,t-1}, c_{i,t-1}, f_{i,t-1}, m_{i,t-1}\right)\right\} = 0. \tag{7}$$

This moment condition highlights the importance of including $G_{i,t-3}$ when identifying the coefficients of the production function. If we omit breeding from the evolution of productivity, that is $\omega_{it} = \phi_t(\omega_{i,t-1}) + \xi_{it}$, then the productivity shock ξ_{it} contains the impact of genetic investment on productivity. If genetic investment correlates with any of these inputs, omitting genetic investment from the law of motion would bias the factor shares. The estimation is conducted in two steps. Using the inputs and the proxy variable, the first stage regression generates predicted output with the estimated input coefficients.[9] Together with the law of motion, the output in the first step is used to construct the productivity shock ξ_{it}. In the final stage, the production function parameters are estimated using the moment conditions mentioned above.

With the method, we address two questions concerning the role of genetic improvement in dairy productivity growth. First, we examine whether farmers' choice in genetics affects the law of motion of productivity. In contrast to previous work, we explicitly model genetic improvement as an investment in future productivity made by the farmer and not solely an exogenous feature of the technology. Second, we examine the extent to which ignoring genetic improvement biases estimated factor shares in the production function when studying dairy farming.

2.2 | Quality adjustment of output

As quality and quantity jointly determine milk price, dairy farmers take both into consideration when making genetic choices. Therefore, we adjust milk production by quality attributes to measure productivity and its growth accurately. Following Atsbeha et al. (2012), we generate a milk quality index with three key attributes affecting milk price: nutrient component percentage, including the butterfat percentage (C_{it}^{fat}) and protein percentage ($C_{it}^{protein}$), and a quality attribute represented by somatic cell count in unit milk (SCC_{it}). The higher the butterfat and protein percentage, the better the milk quality; the lower the SSC, the better the milk quality.

Let I_{it} denotes the unit milk value represented by milk price received by farm i at time t. Assuming that the unit value is approximately linear in protein percentage, fat percentage, and SSC (Atsbeha et al., 2012), we estimate the following hedonic price equation:

---

[9]Ackerberg et al. (2015) discusses the "functional dependence problem" in identifying the coefficients of the labor input β_l in the first stage of the OP and LP methods. The concern is that after controlling investment (or intermediate input) in a nonparametric function for unobserved productivity, there is no variation left to consistently estimate β_l. In our setting, the unique features of the control variable, the investment in breeding and genetics, enable us to break the functional dependence problem and identify the parameters of β_l, β_f, and β_m. As the data generation processes of l_{it}, f_{it}, and m_{it} are different from that of G_{it}, the "optimization errors" (Ackerberg et al., 2015) in the choices of l_{it}, f_{it}, and m_{it} induce variations for identification. Variation also comes from the different timings of the choices of l_{it}, f_{it}, m_{it}, and G_{it}. Whereas l_{it}, f_{it}, and m_{it} are chosen at t as functions of ω_{it}, the choice of G_{it} partially depends on the information of G_{t-3} (Equation (4)). We thank a reviewer for pointing this out.

$$I_{it} = \alpha_0 + \alpha_1 C_{it}^{fat} + \alpha_2 C_{it}^{protein} + \alpha_3 SCC_{it} + \eta_{it}, \tag{8}$$

where the term η_{it} is an error. With the estimated parameters $\widehat{\alpha}_0$, $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, and $\widehat{\alpha}_3$, we calculate the average unit milk value of all sample farms $\widehat{I}$ with average milk quality attributes $\overline{C}^{fat}$, $\overline{C}^{protein}$ and $\overline{SCC}$ as follows:

$$\widehat{\overline{I}} = \widehat{\alpha}_0 + \widehat{\alpha}_1 \overline{C}^{fat} + \widehat{\alpha}_2 \overline{C}^{protein} + \widehat{\alpha}_3 \overline{SCC}. \tag{9}$$

The term $\widehat{I}_{it} / \widehat{\overline{I}}$ is specified as the milk quality index, which equals one when a farm's milk quality is equivalent to the sample average, greater (or lower) than one when a farm's milk quality is better (or lower) than average. The quality-adjusted milk output of farm i, $\widetilde{Q}_{it}$, is then calculated as:

$$\widetilde{Q}_{it} = \frac{\widehat{I}_{it}}{\widehat{\overline{I}}} Q_{it}. \tag{10}$$

2.3 | Survival bias

Firms dropping out of the data present a problem for accurately estimating production functions, especially if firms drop out for having too low of productivity. This is especially a concern when analyzing the dairy sector because there is a long-run trend toward farm consolidation as small farms go out of business and the remaining farms grow larger (MacDonald et al., 2020). If the farms dropping out of our sample have lower productivity than those "surviving," our production function estimates would be biased.

We follow Olley and Pakes (1996) and assume that firm i stays in the market ($\chi_{it} = 1$) instead of exiting the market ($\chi_{it} = 0$) if its productivity exceeds some threshold $\overline{\omega}_{it}(K_{it}, C_{it})$, which makes the sample selection rule:

$$\chi_{it} = \begin{cases} 1 & \omega_{it} \geq \overline{\omega}_{it}(K_{it}, C_{it}), \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Our investment measure G_{it} is also a function of current productivity ω_{it}. The practical implication of this is that the probability of a firm i exiting the market in time t, $P_{it} = P(\chi_{i,t-1} = 0)$, is a function of last year's inputs: $K_{i,t-1}$, $C_{i,t-1}$, and $G_{i,t-1}$. To correct the production function estimates, the probability has to be estimated using these three variables and then included in the final estimation.[10]

Olley and Pakes (1996) use probit and a kernel estimator to estimate these probabilities and include a range of nonlinear interactions of the three variables. We improve on their method by using a supervised machine learning algorithm, random forest, to estimate the probabilities as a function of capital, the number of cows, and genetic investment. The advantage of the random forest is that it naturally takes into account nonlinearities when predicting because it iteratively splits the features in random places to improve model fit. Our random forest model is trained using fivefold cross-validation and the details of the estimation can be found in the supplementary materials in Appendix B. When benchmarked against probit, OLS, and kernel estimation,

---

[10]Details of how P_{it} is incorporated into the estimation are found in Olley and Pakes (1996). Effectively, instead of just approximating a proxy function ϕ_{it} as would be ordinary in a control function approach, another proxy function, ψ_{it}, has to be approximated as a function of $\widehat{\phi}_{i,t-1}$, $K_{i,t-1}$, $C_{i,t-1}$, $G_{i,t-1}$, and the probability estimate $\widehat{P}(K_{i,t-1}, C_{i,t-1}, G_{i,t-1})$. The function ψ_{it} is included in the final estimation of the production function to correct for the sample selection bias.

the random forest performs significantly better even without training. Our analysis in the supplementary materials in Appendix B demonstrates that supervised machine learning can greatly improve this step of estimating production functions that have sample selection. We next turn to our data sources for estimating the production function and understanding the role of genetic investment in dairy farm productivity.

3 | DATA

The main data used in this study consist of 372 farm-level observations of output and inputs from 60 farms in Wisconsin over 2008–2018. The data are collected by the University of Wisconsin Center for Dairy Profitability (CDP). The CDP data are collected by two farm records associations, which provide record-keeping, tax preparation, and consulting services to dairy farms throughout Wisconsin. Compared to the agricultural census, farms that participate in CDP benchmarking are larger: the average in the CDP data is about 177 cows whereas the averages in the 2012 and 2017 census were closer to 120. Given that more than 60% of Wisconsin farms in the census are less than 100 cows, our sample represents slightly larger operations but would still be considered small-scale dairies. In both the CDP and census, less than 10% of farms are larger than 500 cows. Our CDP sample is broadly representative of small- to medium-scale dairies with between 50 and 500 cows.

We also combine the CDP data with records of breeding decisions from Dairy Herd Improvement Associations (DHIAs) over 2012–2018 to measure the genetic quality of bulls each farm selects. DHIAs collect cow-level records for the purpose of benchmarking and genetic evaluation, as well as details on which bulls bred to which cows on the farm. By matching bull identities to their publicly available market traits, we can obtain a measure of genetic quality that is based on which bulls the farm chooses to breed with each year rather than expenditure.

For the dairy farms in the CDP data set, we observe extensive output and input information in each year. On the output side, we observe revenue from selling milk, the quantity of sold milk, butterfat, protein, and the somatic cell counts (SCC) of each unit of milk. On the input side, we observe expenditure on hired labor, feed, fuel, and utility; the number of cows; and the value of buildings, machinery, and equipment owned by the farms. Critically, we also observe breeding expenses, which is expenditure on bull semen and breeding services to produce replacement offspring in the future. To obtain the indirect quantity measures of output and inputs, the revenue and the expenditure are deflated with national level price indices obtained from the USDA Quick Statistics (NASS, 2022).[11],[12] The price indices used to deflate each factor are found in the bottom of Table 1.

Table 1 presents the descriptive statistics of the ouput and inputs, including mean, standard deviation, and percentage change over 2008–2018. During the sample period, the average farm size measured by average milk revenue and herd size shows a growing trend. The average milk revenue increases from 74 million in 2008 to 133 million in 2018 and grows by 80%. The average herd size grew by 64%, from 177 in 2008 to 290 in 2018. During the same period, the average breeding

---

[11]We do not use the observed milk quantity as we want to adjust milk quality as discussed above.

[12]This is the common practice in the literature to obtain quantity measures of output and inputs (De Locker & Syverson, 2021). This generates the so-called revenue-based total factor productivity (TFPR). Given the lack of farm specific output and input prices, aggregate price indices (national level indices in our case) necessarily generate measurement errors in the output and inputs measures, although we focus only on the dairy farms in Wisconsin. One possible solution is to decompose revenue to price and quantity, and then add in a demand structure for price (e.g., De Loecker, 2011). But this solution also requires assumptions on which the literature has not settled. The decomposition method does not work for measurement errors in inputs. We refer the readers to a detailed discussion in Section 4 of De Locker and Syverson (2021). With regard to milk quality, farmers are paid on their components using Class III prices. However, because prices change month to month, there is no way to apply this pricing scheme to milk output measured at the annual level. Also, farms receive individual premiums and incentives that we do not observe. To adjust for milk quality, we use regression-derived weights calculated from Equation (8) to perform the quality adjustment as Atsbeha et al. (2012) does. This procedure necessarily introduces some measurement error in the case that the estimates are inaccurate. Still, we believe this approach is highly preferred to not quality adjusting output, and so the risk of measurment error is justified.

TABLE 1 Summary statistics of CDP sample

Variable	All	2008	2013	2018	Change % over 2008–2018
Milk sold income[1]	91,215 (104,043)	73,788 (86,610)	88,498 (97,403)	132,895 (142,901)	80
Milk (1000 pounds)	5469 (6152)	4347 (5080)	5463 (5992)	7786 (8195)	79
Butter fat (1000 pounds)	202 (224)	161 (186)	202 (214)	297 (312)	84
Protein (1000 pounds)	165 (185)	131 (154)	166 (181)	234 (246)	79
Somatic cell count index (1000 cells/ml)	186 (88)	224 (100)	180 (69)	120 (33)	−46
Herd size	213 (213)	177 (183)	208 (204)	290 (282)	64
Capital (1,000)[2]	376 (462)	326 (523)	357 (398)	516 (527)	58
Feed (1,000)[3]	306 (356)	225 (313)	282 (305)	391 (391)	74
Hired labor (1,000)[4]	117 (154)	105 (146)	122 (156)	136 (187)	30
Utility (100 kWh)	19 (16)	18 (18)	18 (13)	24 (20)	33
Breeding expenditure (1,000)[5]	20 (20)	19 (21)	19 (19)	27 (23)	42
# Observations	372	45	35	19	-

Note: The numbers in the table are means and standard deviations (in parentheses). [1] Milk sold income is adjusted with dairy product price index. [2] The market value of building is adjusted with building material price index. The market value of machinery and equipment is adjusted with machinery price index. [3] The expenditure on feed is adjusted with forage feed price index. [4] The expenditure on hired labor is adjusted with wage rate price index. [5] The breeding expenditure is adjusted with CPI.

expenditure grows by 42%. Table 2 shows the descriptive statistics such as mean, standard deviation, and percent change over 2008–2018 of milk quality indices containing butterfat percentage, protein percentage, and SCC. Table 3 shows the estimates of Equation (8), which we use to construct quality-adjusted output and the "linear adjuster." The linear adjuster is about 0.9, meaning it revises output slightly lower due to farms producing volume without necessarily producing more components. The linear adjustment stays fairly constant as there is only a minor change in butterfat percentage and protein percentage from 2008 to 2018. However, SCC decreased by around 40% during the period. SCC is typically used as a proxy for the incidence of mastitis, so its decrease implies increasing milk quality and cow health.

In this study, we focus on the breeding technology of artificial insemination (AI), which is the most widely used breeding technology in the US dairy sector (Khanal & Gillespie, 2013; Nehring et al., 2021). In AI, semen from selected sires are artificially introduced to cows. AI has wide adoption in the dairy industry because it allows for more precise control over the genetic quality of future replacements and lowers the risk of venereal disease (Foote, 1996). According to data from the Agricultural Resources Managment Survey (ARMS), it was used by more than 80% of dairy farms in 2016 (Nehring et al., 2021). Farms tend to avoid buying replacements from outside the farm due to the higher cost and the risk of disease. AI is generally the preferred method for choosing the future genetics of the herd in Wisconsin.

Figure 2 shows the distribution of breeding expenditure, our main investment proxy, across farms in our whole sample. Breeding expenditure is made up of two components: genetic quality

T A B L E 2   Summary statistics of milk quality indexes

Symbol	Description	Unit	All	2008	2013	2018	Change % over 2008–2018
C_{it}^{fat}	Butterfat	%	3.75 (0.24)	3.76 (0.26)	3.80 (0.19)	3.84 (0.14)	2
$C_{it}^{protein}$	Protein	%	3.04 (0.19)	3.09 (0.37)	3.06 (0.11)	3.04 (0.08)	−2
SCC_{it}	Somatic cell count index	1000 cells/ml	200 (90)	218 (93)	180 (70)	134 (40)	−39
$\frac{\widehat{I_{it}}}{\widehat{I}}$	Linear Adjuster	#	0.91 (0.03)	0.92 (0.03)	0.88 (0.02)	0.90 (0.01)	−2
Q_{it}	Milk production	1000 pounds	4514 (5448)	4087 (4751)	4707 (5504)	7334 (7937)	79
$\widetilde{Q}_{it}$	Adjusted milk production	-	4080 (4916)	3791 (4436)	4120 (4786)	6565 (7093)	73
	# Observations		738	60	46	28	

Note: The numbers in the table are means and standard deviations are in parentheses.
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

T A B L E 3   Quality adjustment parameters

Variable	Parameters
Butterfat percentage	1.057*** (0.186)
Protein percentage	0.572** (0.244)
Somatic cell count index	−0.003*** (0.001)
R-square	
Within herd	0.3188
Between herd	0.6161
Overall	0.4138
# Observation	738
# Herds	116

Note: The parameters are estimated using OLS with the farm IDs as fixed effects.
***$p < 0.01$, **$p < 0.05$.

and technician fees. Technician fees are usually more or less uniform across cows, but the remaining variation in expenditure is due to the cost of the genetics itself. Hedonic analyses of dairy genetics markets show that the most expensive genetics tend to be those with high production traits and, to a lesser extent, bulls with better health traits (Richards & Jeffrey, 1996; Schroeder et al., 1992).

    To adjust for herd size, Figure 2 shows breeding expenditure per cow. On average, farms spend about $90 per cow. As a reference point, the average price for a dairy bull was $20–$30 per service in the period 2010–2020. Because some cows do not get pregnant the first time, we may expect that in 1 year a farmer may purchase two bull services, meaning an annual cost between $40 and $60 per cow per year. The remaining cost can be explained by additional service costs charged by the breeding company, which do not scale with genetic quality in the majority of cases.[13] The median breeding expenditure is $85 due to the large right tail of farms that spend more than $150 per cow on

---

[13]Large outliers may be explained by embryo transfer, a more expensive delivery method for genetics used in a minority of cases.
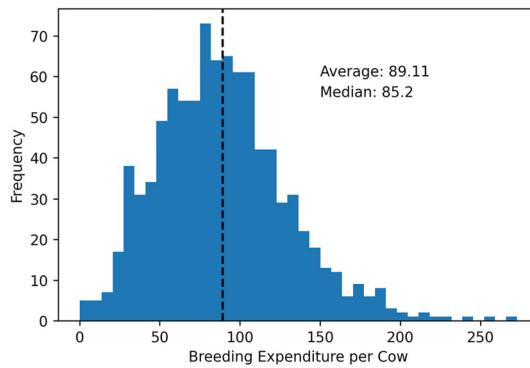
FIGURE 2   The distribution of breeding expenditure per cow

breeding. Although not much data are available on the costs of individual genetics on dairy farms in general, the price of bulls is primarily explained by genetic quality (Richards & Jeffrey, 1996). For this reason, the breeding expenditure variable is a suitable proxy for investment in future productivity through genetic improvement.

One potential component of breeding expenditure that can attenuate its impact on productivity is repeated breeding attempts. If a cow does not conceive after the farm pays for breeding services, the farm must purchase the genetics again to repeat the attempt. To the extent that some infertility may be a result of management, breeding expenditure may be high in some cases because of persistently poor fertility. If firms with poor fertility are also less productive, then this attenuates the positive impact of breeding expenditure on productivity toward zero. We should also expect that farms with high productivity get more productivity out of their breeding expenditure, meaning the interaction between productivity and breeding expenditure (θ_2 in Equation (6)) would be positive.

To complement our analysis using breeding expenditure, we merge CDP data with our DHIA data to obtain a more direct measure of genetic investments: the genetic traits of the bulls that farms choose. Dairy farms make investments into the genetics of their herd by purchasing bulls with certain traits. Farms can choose bulls with higher production traits, for example a higher fat or protein yield, in order to have replacement cows in the future that yield more fat or protein. Fat and protein yield are especially important to Wisconsin dairy farms because they are located in the Upper Midwest Milk Marketing Order (Federal Milk Marketing Order 30) and are paid based on the amount of protein and fat they produce in pounds. Given these incentives, most dairy bull selection indices do not include milk yield independent of fat and protein (that is increasing volume). Instead, fat and protein yield are the primary trait indices using in dairy bull selection.[14] Farms can also choose to invest in health by choosing bulls with a lower "somatic cell count" (a measure that correlates to the incidence of mastitis) or a higher "daughter pregnancy rate" (a measure that correlated to fertility).

Our data on genetic indices come from the Council on Dairy Cattle Breeding (CDCB). The CDCB estimates genetic indices for dairy sires called "predicted transmitting ability" (PTAs) three times a year, which are publicly available through the National Association for Animal Breeders (NAAB, 2022). The indices measure how much a sire will "transmit" performance relative to a base bull whose PTA is equal to zero. For example, if a bull has a butterfat PTA score of 50, then the farmer can expect that offspring from that bull will produce 50 more pounds of butterfat than the base bull. The indices are used by companies to price bulls and by dairy farmers to help inform their breeding choices.

---

[14]See, for example, the weights for the Net Merit index used widely in the dairy industry. In this index, fat and protein are valued at $4 a pound whereas milk yield is valued at less than $0.01 a pound. Milk yield is given a low weight because, independent of fat and protein yield, there is little market value to producing high volume.

We use around 140,000 dairy bull evaluations from the period 2012–2019 and focus on five traits: milk yield, fat yield, protein yield, somatic cell count, and daughter pregnancy rate. The first three are production related, the fourth is related to health, and the last trait is a measure of fertility. To construct production and health indices from these five traits, we use a technique called principal component analysis (PCA). PCA is an orthogonal, linear transformation of data into a new coordinate system whereby the first axis, called the "first component," explains the most variance in the data possible, the second axis the second most variance, and so on. PCA was introduced to the economics and psychology literature by Hotelling (1933) and has since been applied in economics in a variety of ways (Aït-Sahalia & Xiu, 2019; Manyong et al., 2006; Nieuwoudt, 1972). One popular and relevant application of PCA is constructing indices of several different variables (Vyas & Kumaranayake, 2006). Because each component is a linear combination of variables in the data, the components can be interpreted as a weighted index of the variables. This is especially useful when there is correlation between different variables, and we would prefer one index that represents them all.

In the case of genetic traits, PCA helps us construct two indices that represent our three production traits and two health traits. In our calculations, the first two components captured 83% of the variance in these five traits. The first component was highly correlated with the three production traits, whereas the second component was highly correlated with somatic cell score and daughter pregnancy rate. We use the first component as our "production index" and the second as our "health index."[15]

After constructing these indices, we match this data to DHIA data of farm breeding choices. By matching the bull ID in DHIA to its evaluation from the CDCB, we know the bull's production and health index scores at the time it was chosen. This gives us two measures of genetic investment: breeding expenditure and genetic indices. The first of these measures is simply a cost measure, how much dairy farms spend on breeding, whereas the second more accurately measures in *what* dairy farms are investing. Unfortunately, there are only a few farms both in DHIA and CDP so the genetic index scores can only be used in a subset of our data. Still, by examining the subset of data belonging to both the CDP data and the DHIA, we have a never-before-seen look into how dairy farmer breeding choices translate into productivity.

The summary statistics for this merged sample are in Table 4. The merged CDP-DHIA sample has 88 observations over 2012–2018. The average herd size in the merged sample is 275 cows per farm, larger than the average herd size in the CDP sample, 213 cows per farm. The average output and input are also larger for dairy farms in the merged CDP-DHIA sample.

Combining both CDP and DHIA data, we obtain two proxy measures for genetic investments: breeding expenditure and the average production and health indices of their bull choices. These two measures capture different aspects of genetics investments. Total breeding expenditure contains the cost of the bull semen, labor fees, and management fees. According to hedonic analyses of dairy bull prices, bulls with the most productive genetics in terms of milk production tend to be the most expensive, so we can assume breeding expenditure is weakly increasing in genetic traits measuring productivity (Richards & Jeffrey, 1996; Schroeder et al., 1992). Our genetic indices measure the average traits chosen in that year, which helps us measure the type of investments being made. This measure also partially alleviates concerns about potential distortion on total genetic investment caused by herd size. It is not a per-cow based measure as it is averaged over chosen bulls. However, it reflects the average quality of genetics incorporated into a herd and should be highly correlated with the average level of investment, especially for large herds. We analyze both production traits and health traits using our created indices. Milk producing traits have received the most attention (Atsbeha et al., 2012; Roibas & Alvarez, 2010), yet health traits can also have an important relationship to productivity. Many health conditions such as mastitis directly impact milk production, making it feasible that farms may invest in these health traits to increase productivity.

---

[15]Table A1 in the supplementary appendix reports the loading scores for the production index and shows that milk yield, fat yield, and protein yield are correlated to the first component in a similar way. This indicates there is no direct trade-off between milk yield and component yield. We refer the readers to the supplementary Appendix A for a more detailed discussion.

TABLE 4 Summary statistics for CDP-DHIA sample

Variable	Mean	Std. err.	Minimum	Maximum
Milk sold income (1000)[1]	127,665	137,428	10,140	632,094
Milk (1000 pounds)	7469	7867	640	35,528
Butter fat (1000 pounds)	279	288	24	1305
Protein (1000 pounds)	225	236	20	1062
Somatic cell count index	148	58	61	374
Herd size	275	259	38	1177
Capital (1000)[2]	517	536	4	2401
Feed (1000)[3]	424	446	13	2150
Hired labor (1000)[4]	147	191	0	855
Utility (100 kWh)	23	19	4	84
Breeding expenditure (1000)[5]	26	24	1	92
Genetic production index	0.80	0.56	−1.30	1.91
Genetic health index	0.30	0.44	−1.32	1.40
# Observations	88			

Note: [1] Milk sold income is adjusted with dairy product price index. [2] The market value of building is adjusted with building material price index. The market value of machinery and equipment is adjusted with machinery price index. [3] The expenditure on feed is adjusted with forage feed price index. [4] The expenditure on hired labor is adjusted with wage rate price index. [5] The breeding expenditure is adjusted with CPI.

Figure 3 shows the growing trend of average breeding investment and the production and health ability of the chosen sires in our sample. These Wisconsin dairy farms made breeding decisions similar to the rest of the country: The genetic production index is growing rapidly over time just as in Figure 1. In contrast, the health index scarcely grew at all during this period. There seems to be little variation in the health index during this period, reflecting a tendency for dairy farmers to invest in milk production more than in health traits.

In the next section, we present our estimation results from this unique dataset to explore the role of breeding in accelerating productivity growth in dairy. Comparing across methods, we also determine whether omitting genetic investment leads to biased estimates of factor shares.

4 | RESULTS

Panel A of Table 5 presents the parameters estimated with four models: an OLS model, a standard LP model in which breeding expenditure is the proxy variable for unobserved productivity (but not in the law of motion), a LP model with breeding expenditure in the law of motion (Equation (5)), and an LP model where breeding expenditure is interacted with productivity in the law of motion (Equation (6)). In these last two models, breeding expenditure is also used as the proxy variable. The parameters of interest in LP models are estimated in the GMM framework, and one-period lag terms are used as the instruments for labor, feed, and intermediate inputs. The last column of Table 5 reports the means and standard errors of the differences of the estimated input coefficients in Columns 2 and 4.

Comparing the results in Column 1 with those of the other columns, we find that the OLS overestimates the coefficients of the flexible inputs of labor and herd size represented by cow number because of the endogeneity or the positive relation between the inputs and unobserved productivity. This is consistent with the findings in the literature (e.g., Jang & Du, 2019).[16] Including breeding

---

[16]We do not find a downward bias of the OLS estimate on the capital input as suggested in the literature probably because of relatively small size and number of farms in the sample. We thank a reviewer for pointing this out.
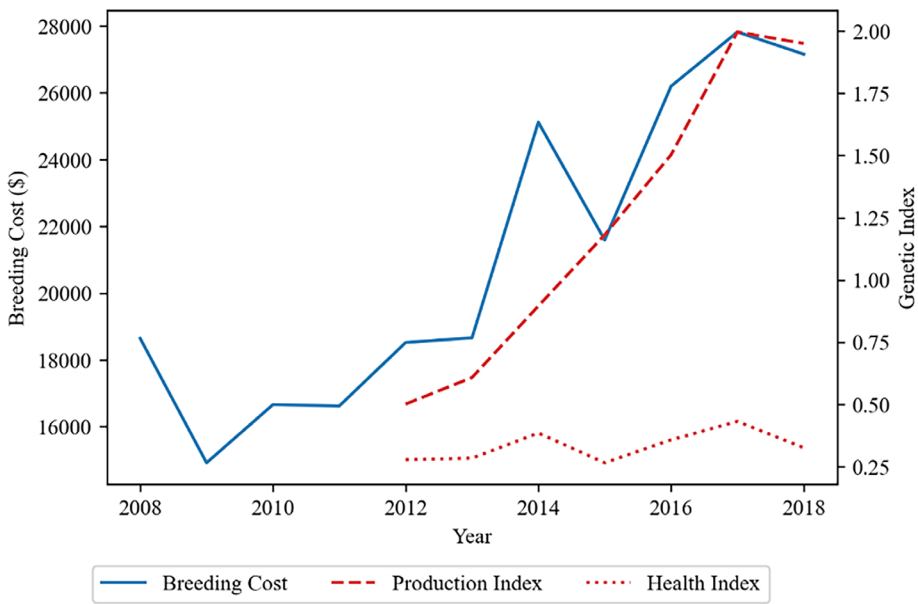
FIGURE 3   Average breeding investment and genetic production index, 2008–2018

investment in the law of motion of productivity has no statistically significant effects on the estimated factor shares. Coefficients do change, but the differences between Column 2 and 4 are not statistically significant. Including a nonlinear interaction term does not appear to affect estimation given that the Column 3 and Column 4 estimates are essentially the same. Although including breeding investments may affect factor shares, our sample size is too small to determine whether these changes are statistically significant.

Panel B of Table 5 presents the coefficients for the law of motion of productivity. As expected, breeding expenditure that occurred 3 years ago has a positive and significant effect on current productivity. Although there was some risk of attenuation bias using breeding expenditure, we still see a positive and significant impact. The probability of exiting is negatively related to productivity, which implies that firms more likely to exit have lower productivity. This fits with what we expect from an industry that is consolidating: less productive firms are exiting and leaving only the most productive firms in the industry (Olley & Pakes, 1996). When we include the interaction term, we see a negative sign on the interaction, but only the probability of exit remains statistically significant. This pattern suggests that there are diminishing returns to investments in genetics, but the size of the standard errors makes it impossible to confirm whether this is the case. Still, because we do not find the interaction term to be positive, this is evidence against breeding attempts being a major component of breeding expenditure (see discussion in Section 3).

Table 6 examines the same questions but compares using breeding expenditures to genetic indices, a less noisy measure of investment in genetic improvement. Panel A of Table 6 compares the factor shares for the models that use breeding investment, genetic indices, or both in the law of motion respectively. The coefficients change very little between these models, except the materials coefficients that are not statistically different from zero in any model. Panel B of Table 6 shows how the genetic indices factor in the law of motion compare to that of breeding expenditure. Production and health indices from bulls chosen 3 years ago correlate positively to current productivity, though only the production index is statistically different than zero at the 95% level. The interaction term for the production index is negative just as in breeding expenditure and is statistically significant.

T A B L E 5 Estimates of the production function and Law of Motion

	OLS	Standard LP model	De Loecker model		Coefficient difference, (2)–(4)
	Panel A: Production Function				
Variable	(1)	(2)	(3)	(4)	
Labor	0.077*** (0.012)	0.067*** (0.012)	0.059 (0.047)	0.059*** (0.010)	0.008 (0.006)
Capital	0.022*** (0.008)	0.016** (0.007)	0.016** (0.007)	0.017*** (0.005)	−0.001 (0.004)
Herd Size	0.809*** (0.026)	0.786*** (0.052)	0.795*** (0.048)	0.804*** (0.038)	−0.018 (0.036)
Feed	0.110*** (0.021)	0.111* (0.062)	0.116** (0.048)	0.105*** (0.032)	0.005 (0.039)
Material	0.026 (0.020)	0.079* (0.044)	0.041 (0.073)	0.041*** (0.029)	0.038 (0.027)
Constant	8.233*** (0.143)	8.062*** (0.045)	8.053*** (0.034)	7.994*** (0.038)	0.068 (0.056)
# Observations	366	366	366	366	
Interaction	-	-	No	Yes	-
	Panel B: Law of Motion				
Productivity, $t-1$	-	-	0.523*** (0.071)	0.614 (0.383)	-
Breeding Expend, $t-3$	-	-	0.014*** (0.003)	0.117 (0.316)	-
Productivity × Breeding Expend			-	−0.012 (0.038)	
Exit Probability $P_t(k_{t-1}, b_{t-1})$	-	-	−0.060*** (0.021)	−0.062*** (0.022)	-
# Observations	-	-	283	283	-
Year FE	-	-	Yes	Yes	-
R-squared	-	-	0.726	0.723	-

Note: Standard errors are in parentheses.
***$p < 0.01$,**$p < 0.05$,*$p < 0.1$.

The interaction term is an estimate of the parameter θ_2 from Equation (6), which measures how the new cohort's improved productivity enhances the productivity of the current cohort. We find evidence of diminishing returns to genetics, meaning already productive farms will benefit less from genetic investments than farms that have less productive cows. This makes intuitive sense, because cows are likely to have biological constraints that prevent genetic improvement from indefinitely increasing productivity. For example, cows that have higher milk production are more likely to develop certain health problems (Oltenacu et al., 2010). Farms with lower productivity have more to gain from genetic improvement than farms that are already productive and may be dealing with this trade-off between health and production.

As a robustness check, we examine the relationship between realized quality and genetic investments by comparing the coefficients on the law of motion with and without quality-adjusted output. Because many genetics investments involve quality (e.g., protein and butterfat output), it is important to understand how quality and genetics investments interact. Table 7 show the model parameters from Table 6 with and without the quality adjustment and the difference between them. As before, the production function coefficients differ but are not statistically different, which is likely due to small sample size. In

TABLE 6   Production function, CDP-DHIA sample

Variable	Breeding expenditure	Genetic indices
	Panel A: Production function	
	(1)	**(2)**
Labor	0.057***	0.055
	(0.019)	(0.107)
Capital	0.013	0.011
	(0.009)	(0.029)
Herd size	0.819***	0.817***
	(0.054)	(0.243)
Feed	0.124*	0.147**
	(0.069)	(0.062)
Material	0.009	0.031
	(0.048)	(0.081)
Constant	8.025***	8.082***
	(0.049)	(0.162)
Observations	87	87
R-squared	0.609	0.491
	Panel B: Law of Motion	
Productivity, t−1	0.664	0.970***
	(1.421)	(0.340)
Breeding expend, t−3	0.296	-
	(1.153)	
Production index, t−3	-	3.895**
		(1.645)
Health index, t−3	-	2.873
		(2.718)
Productivity, t−1 × Breeding expend, t−3	−0.033	-
	(0.135)	
Productivity, t−1 × Production index, t−3	-	−0.482**
		(0.203)
Productivity, t−1 × Health index, t−3	-	−0.351
		(0.335)
Exit probability, $P_t(k_{t-1}, b_{t-1})$	−0.185**	−0.186**
	(0.070)	(0.082)
# Observations	55	55
Year FE	Yes	Yes
R-squared	0.609	0.491

***$p < 0.01$,**$p < 0.05$,*$p < 0.1$.

the unadjusted model, more of the coefficients such as labor and feed are insignificant than in the quality-adjusted model. When the quality dimension is not taken into account, it appears that many inputs have no relationship to output. Similarly, investments in production traits have no impact on productivity in the law of motion when output is not quality adjusted.

There are two insights generated from this exercise. First, inputs may be used to generate quality (butterfat and protein) instead of volume. If this is the case, ignoring quality adjustment may lead us to think some inputs do not matter to production. The second insight is that farms investing in quality are likely also making breeding investments into quality. Without adjusting for quality, these investments will not appear to have any impact on productivity.

TABLE 7 Model results with and without quality adjustment

Variable	Unadjusted model	Quality-adjusted model	Difference between two models
	Panel A: Production function		
Labor	0.056	0.055	0.001
	(0.059)	(0.107)	(0.073)
Capital	0.000	0.011	−0.011
	(0.017)	(0.029)	(0.025)
Herd size	0.802***	0.817***	−0.016
	(0.149)	(0.243)	(0.265)
Feed	0.153	0.147**	0.006
	(0.093)	(0.062)	(0.148)
Material	0.050	0.031	0.019
	(0.138)	(0.081)	(0.100)
Constant	8.009***	8.082***	−0.074
	(0.091)	(0.162)	(0.138)
Observations	87	87	-
	Panel B: Law of Motion		
Productivity, t−1	0.726**	0.970***	0.244
	(0.326)	(0.340)	(0.471)
Production index, t−3	1.195	3.895**	2.700
	(1.917)	(1.645)	(2.526)
Health index, t−3	4.094	2.873	−1.220
	(2.892)	(2.718)	(3.968)
Productivity, t−1 × Production index, t−3	−0.147	−0.482**	−0.336
	(0.235)	(0.203)	(0.310)
Productivity, t−1 × Health index, t−3	−0.503	−0.351	0.152
	(0.355)	(0.335)	(0.488)
Exit probability, $P_t(k_{t-1}, b_{t-1})$	−0.260***	−0.186**	0.073
	(0.095)	(0.082)	(0.125)
# Observations	55	55	-
Year FE	Yes	Yes	-
R-squared	0.490	0.491	-

Note: Standard errors are in parentheses.
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

5 | CONCLUSION

Our work examines the relationship between investments in genetics through breeding and productivity growth in the US dairy industry. Using a detailed dataset of Wisconsin dairy farms, we incorporate investment in breeding services and genetic indices into the productivity law of motion. We test whether or not omitting genetics investments impacts the estimation of factor shares in the dairy production function. We find that including genetics investments in the law of motion changes the point estimates for feed, labor, and materials in the production function, though no differences are statistically different than zero. Although breeding investments made 3 years ago positively impact productivity in the current period, the returns to investment are lowest for farms with high productivity. Put differently, we find that there are diminishing returns to investments in genetics on Wisconsin dairy farms.

Our work has several limitations that may be able to be addressed in future work. First, the sample size in this study is very small. The sorts of data needed to understand investments in genetics in the context of other inputs are hard to come by. Although datasets like the Agricultural Resource Management Survey (ARMS) have detailed information about capital and labor inputs, they do not have information about breeding investments. This a critical data gap that still exists when studying dairy farm operations and leads us to our second limitation. In this study, we use spending on breeding as a proxy for genetic investments. This may be a crude proxy for what we are actually trying to measure: investments in genetic quality. Data on individual breeding decisions from DHIA can help fill this gap, but the lack of crossover in our sample between DHIA and CDP farms makes estimation very imprecise. Also, because the breeding decisions are at the animal level, it is an open question what the best way to aggregate this information to the farm level is. Further data collection efforts should particularly focus on trying to understand how much dairy farmers pay for different kinds of genetics. These kinds of data are vital to understanding the link between investments in genetics and productivity growth in a more precise way.

Strategic breeding has been one of the most important innovations in animal agriculture in the past century. The next century poses new challenges for animal industries such as dairy, especially addressing excessive greenhouse gas emissions from livestock industries. Part of these externalities may be addressed in the future by breeding away from production and toward sustainability of the operation (Wall et al., 2010). Our work is a crucial first step in determining just how important farm investments in genetic improvement have been in the on-farm dynamics of productivity growth, and thus a crucial first step in better understanding the role of genetics in the future of the dairy industry.

REFERENCES
Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica* 83(6): 2411–51.

Aït-Sahalia, Yacine, and Dacheng Xiu. 2019. "Principal Component Analysis of High-Frequency Data." *Journal of the American Statistical Association* 114(525): 287–303.

Atsbeha, Daniel M., Dadi Kristofersson, and Kyrrel Rickertsen. 2012. "Animal Breeding and Productivity Growth of Dairy Farms." *American Journal of Agricultural Economics* 94(4): 996–1012.

Bravo-Ureta, Boris E., Alan Wall, and Florian Neubauer. 2021. "Dairy Farming from a Production Economics Perspective: An Overview of the Literature." In *Handbook of Production Economics*, edited by S. C. Ray, R. G. Chambers, and S. C. Kumbhakar, 1–39. Singapore: Springer Singapore.

De Locker, Jan, and Chad Syverson. 2021. "*Handbook of Industrial Organization*." In *An Industrial Organization Perspective on Productivity*, Vol IV, edited by Kate Ho, Ali Hortaçsu, and Alessandro Lizzeri 141–223. North-Holland: Elsevier.

De Loecker, Jan. 2011. "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberation on Productivity." *Econometrica* 79(5): 1407–51.

De Loecker, Jan. 2013. "Detecting Learning by Exporting." *American Economic Journal: Microeconomics* 5(3): 1–21.

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and Firm-Level Export Status." *American Economic Review* 102(6): 2437–71.

Foote, Robert H. 1996. "Dairy Cattle Reproductive Physiology Research and Management—Past Progress and Future Prospects." *Journal of Dairy Science* 79(6): 980–90.

Frick, Fabian, and Johannes Sauer. 2018. "Deregulation and Productivity: Empirical Evidence on Dairy Production." *American Journal of Agricultural Economics* 100(1): 354–78.

Grieco, Paul L., Shengyu Li, and Hongsong Zhang. 2016. "Production Function Estimation with Unobserved Input Price Dispersion." *International Economic Review* 57(2): 665–90.

Griliches, Zvi, and Jacques Mairesse. 1998. "*Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium.*" In *Production Functions: The Search for Identification.* Cambridge, UK: Cambridge University Press.

Hall, Robert. 1988. "The Relation between Price and Marginal Cost in U.S. Industry." *Journal of Political Economy* 96: 921–47.

Hoch, Irving. 1955. "Estimation of Production Function Parameters and Testing for Efficiency." *Econometrica* 23: 325–6.

Hotelling, Harold. 1933. "Analysis of a Complex of Statistical Variables into Principal Components." *Journal of Educational Psychology* 24(6): 417–41.

Hutchins, Jared, and Brent Hueth. 2021. "Supply Response in Dairy Farming: Evidence from Monthly, Animal-Level Data." *Journal of Agricultural and Resource Economics* 47(1): 38–56.

Hutchins, Jared, Brent Hueth, and Guilherme Rosa. 2021. "Quantifying Heterogeneous Returns to Genetic Selection: Evidence from Wisconsin Dairies." In *Economics of Research and Innovation in Agriculture*, edited by P. Moser, 81–106. Chicago: University of Chicago Press.

Jang, Heesun, and Xiaodong Du. 2019. "Evolving Techniques in Production Function Identification Illustrated in the Case of the us Dairy." *Applied Economics* 51(14): 1463–77.

Khanal, Aditya, and Jeffrey Gillespie. 2013. "Adoption and Profitability of Breeding Technologies: Evidence from us Dairy Farms." *AgBioforum* 16(1): 53–65.

Kirwan, Barrett E., Shinsuke Uchida, and T. Kirk White. 2012. "Aggregate and Farm-Level Productivity Growth in Tobacco: Before and after the Quota Buyout." *American Journal of Agricultural Economics* 94(4): 838–53.

Läpple, Doris, Colin A. Carter, and Cathal Buckley. 2021. "EU Milk Quota Abolition, Dairy Expansion, and Greenhouse Gas Emissions." *Agricultural Economics* 53(1): 125–42.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies* 70(2): 317–41.

MacDonald, James M., Jonathan Law, and Roberto Mosheim. 2020. *Consolidation in US Dairy Farming*. Kansas City: Economic Research Service, USDA.

Manyong, Victor M., I. Okike, and Timothy O. Williams. 2006. "Effective Dimensionality and Factors Affecting Crop-Livestock Integration in West African Savannas: A Combination of Principal Component Analysis and Tobit Approaches." *Agricultural Economics* 35(2): 145–55.

Mukherjee, Deep, Boris E. Bravo-Ureta, and Albert De Vries. 2013. "Dairy Productivity and Climatic Conditions: Econometric Evidence from South-Eastern United States." *Australian Journal of Agricultural and Resource Economics* 57(1): 123–40.

National Agricultural Statistics Service (NASS). (2022). "Quick Stats." https://quickstats.nass.usda.gov/.

National Association for Animal Breeders (NAAB). (2022). "Database files." https://www.naab-css.org/ database-files.

Nehring, Richard F., Jeffrey Gillespie, Catherine Greene, and Jonathan Law. 2021. "The Economics and Productivity of Organic Versus Conventional us Dairy Farms." *Journal of Agricultural and Applied Economics* 53(1): 134–52.

Nieuwoudt, W. L. 1972. "A Principle Component Analysis of Inputs in a Production Function." *Journal of Agricultural Economics* 23(3): 277–83.

Njuki, Eric, Bravo E. Bravo-Ureta, and Victor E. Cabrera. 2020. "Climatic Effects and Total Factor Productivity: Econometric Evidence for Wisconsin Dairy Farms." *European Review of Agricultural Economics* 47(3): 1276–301.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64(4): 1263–97.

Oltenacu, Pascal A., and Donald M. Broom. 2010. "The Impact of Genetic Selection for Increased Milk Yield on the Welfare of Dairy Cows." *Animal Welfare* 19(1): 39–49.

Pryce, J. E., and R. F. Veerkamp. 2001. "The Incorporation of Fertility Indices in Genetic Improvement Programmes." *BSAP Occasional Publication* 26(1): 237–49.

Richards, Timothy J., and Scott R. Jeffrey. 1996. "Establishing Indices of Genetic Merit Using Hedonic Pricing: An Application to Dairy Bulls in Alberta." *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 44(3): 251–64.

Roibas, David, and Antonio Alvarez. 2010. "Impact of Genetic Progress on the Profits of Dairy Farmers." *Journal of Dairy Science* 93(9): 4366–73.

Schroeder, Ted C., Juan A. Espinosa, and Barry K. Goodwin. 1992. "The Value of Genetic Traits in Purebred Dairy Bull Services." *Applied Economic Perspectives and Policy* 14(2): 215–26.

Shook, G. E. 2006. "Major Advances in Determining Appropriate Selection Goals." *Journal of Dairy Science* 89(4): 1349–61.

Townsend, Robert, and Colin Thirtle. 2001. "Is Livestock Research Unproductive? Separating Health Maintenance from Improvement Research." *Agricultural Economics* 25(2–3): 177–89.

Van Raden, Paul. 2004. "Invited Review: Selection on Net Merit to Improve Lifetime Profit." *Journal of Dairy Science* 87(10): 3125–31.

Vyas, Seema, and Lilani Kumaranayake. 2006. "Constructing Socio-Economic Status Indices: How to Use Principal Components Analysis." *Health Policy and Planning* 21(6): 459–68.

Wall, E., G. Simm, and D. Moran. 2010. "Developing Breeding Schemes to Assist Mitigation of Greenhouse Gas Emissions." *Animal* 4(3): 366–76.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hutchins, Jared, Yating Gong, and Xiaodong Du. 2023. "The Role of Animal Breeding in Productivity Growth: Evidence from Wisconsin Dairy Farms." *American Journal of Agricultural Economics* 1–20. https://doi.org/10.1111/ajae.12374