Information Provision and Network Externalities: The Impact of Genomic Testing on the Dairy Industry

Victor Funes-Leal[1] and Jared Hutchins[2]

[1]*Department of Agricultural and Consumer Economics, University of Illinois*

[2]*Department of Agricultural and Consumer Economics, University of Illinois*

Abstract

We develop a synthetic control group to estimate the long-term impacts of introducing Genomic Selection in the US market for dairy cattle genetics. Using a data set of all bulls' genetics marketed in the US and Canada from 2000 to 2020, we find that genomic selection significantly increased genetic gains for all measured traits, particularly production such as milk, protein, and fat yields, but it is also the source of increased levels of **inbreeding depression**, a decrease in the performance of animals whose parents have a high degree of relatedness, as a consequence of genetics companies breeding more animals from established lines rather than lesser-known lines with lower trait values but less inbred. We argue that this is a **network externality** inflicted by genetics companies upon themselves. Our estimation shows that the increased inbreeding rate of American bulls caused a loss of between 2 and 5 billion dollars per year to the entire industry. Solving this externality will require either a mechanism to internalize the harmful effects, such as paying a much higher price for more inbred sires or a collective action mechanism to select which lines will be bred in the next generation.

1 Introduction

Improved genetics via the selection of bulls (sires) and cows (dams) is a crucial driver for productivity improvements in the dairy industry. US milk production has tripled in the last forty years, and over half of this increase is solely due to genetics. The United States is also the world's largest exporter of bovine genetics, with a share of 46.4% of the total value of exports in 2019 (Canada is the second largest with a 19.3% share); every improvement in genetics that takes place in the United States will be quickly transmitted to the rest of the world dairy industry.

The demand for cattle genetics is made up of all dairy farmers in the US and abroad who periodically buy bull semen to impregnate their cows for two reasons: a cow needs to be pregnant or weaning her calf to produce milk, and also because they are going to need new cows in the upcoming years; a cow's productive life is about four years. Dairy farmers rarely keep bulls within their herds because they do not contribute to their profits, so genetics companies provide the semen to impregnate cows. On the other hand, the supply of dairy genetics is constituted by genetics companies that breed bulls to meet the demand for specific traits. To do so, they choose sires and dams with a particular objective. By doing so, they ensure that the next generation of bulls will have a similar or higher level of traits.

Genomic selection uses genetic markers, DNA sequences with a specific location in a chromosome, that cover the entire genome to detect gene combinations whose distribution is significantly different than the expected frequency if they were randomly associated. The most salient consequence of this new method is that it is now possible to test an animal for the presence of markers (which are correlated with specific traits) as soon as the animal is born. What are the consequences of genomic selection? First, it increases the rate of genetic change, that is, the intergenerational variation in the quantities of any genetic trait; this rate of change is a function of how choosy breeders are and how long they must wait for a bull to start fathering his progeny.

The process of gathering measures of observed traits from an animal's progeny and using them to estimate their contribution to the characteristics of his descendants is called evaluation, and it is typically carried out three times a year by the Council on Dairy Cattle Breeding (CDCB)[1]. Each trait is a measure of a specific characteristic of the animal's productivity (milk, fat, protein yield); health (stature, weight, size of udders); or reproduction (calving ability, metritis, gestation length) that a bull will transmit to his offspring, after controlling for factors such as breed, age and, pedigree. In other words, they represent the purely genetic component of the expression of

---

[1] https://uscdcb.com/

any given characteristic, and that is what a cow fathered by a specific bull will, on average, inherit from him.

In this article, we argue that the introduction of a new method to identify traits in dairy cattle by testing their genome has led to faster gains in production and health traits but has also increased the prevalence of inbreeding depression, a decrease in observed trait values attributable to a higher frequency of recessive genes due to the high relatedness of an animal's parents. We argue that this results from genetics companies supplying more bulls from "prestigious," more inbred pedigrees rather than lesser-known but less inbred alternative pedigrees. Companies do so because there is a demand for high-performance bulls, and their objective is to supply the best possible animals. Unfortunately, they do not consider that the average inbreeding rate across all animals has steadily increased since 2010, and they are now beginning to experience its negative impacts (Cole, 2019).

In economic terms, inbreeding depression can be understood as a negative externality that genetics companies inflict on one another because of genetics companies' increased selection of inbred lines. In this case, the profits of genetics companies are a function of the trait levels of the lines they breed but also depend on the degree of relatedness of such lines to each other. The more genetic companies adopt a line, the higher the likelihood of observing inbreeding depression in the next generation.

This is an example of a **network externality** (Liebowitz and Margolis, 1994), (Katz and Shapiro, 1985), where companies enjoy the benefits (higher trait levels) and suffer the costs (lower fertility levels) from changes in the size of their associated network. In this case, the pedigree (relatedness) is the force that binds all the nodes in the network, where two animals are part of the same network (line) if they descend from the same ancestor.

The more genetics companies breed the same lines, the more newborn bulls will be related to a specific ancestor and the denser their networks will be. We no longer have a measure of demand strength since the advent of genomic testing, as bull's traits are no longer measured through their daughters. We can, however, approximate it using the number of sons, grandsons, and the rest of his male progeny; a bull with a high level of traits is likely to have a large number of descendants.

Our main finding is that introducing genomic selection has increased productivity gains. For example, the Total Productivity Index, a weighted average of fourteen genetic traits used to select the best bulls, increased twice in 2009-2019, compared to more moderate growth in 2000-2009 (Figure 2.) On the other hand, it has also led to a higher inbreeding among Holstein cattle and,

with it, a greater prevalence of inbreeding depression, a negative (indirect) network externality inflicted by genetics companies on themselves by "doubling down" on more famous lines in detriment of less inbred lines. We expect that the cost of inbreeding depression will continue to rise in the following years, but the problem cannot be solved by just breeding fewer inbred bulls; it will also require a mechanism to take into account the inbreeding level of the entire Holstein herd.

2   Conceptual Framework

2.1   The Economics of Selective Breeding

In any industry that involves animal products, such as beef, dairy, or poultry, selective breeding is the oldest and most ubiquitous means to increase productivity (Kerr, 1984). Breeding has been practiced since the domestication of cattle, but it took off in the first half of the twentieth century primarily due to a series of progresses in genetics (Lush, 1937), which allowed breeders to identify the link between desired traits in animals and their inheritance mechanisms[2].

The supply chain of the dairy genetics markets comprises two parts; the first is animal breeders, whose objective is to breed the best bulls (those with the highest breeding values). The second part is the genetics companies that sell the semen from those bulls to commercial dairy farmers, who choose which animal to buy based on his trait levels and pedigree.

Genetic improvement occurs only in the first part, where bulls and cows are selected to breed the next generation of animals. The most common method breeding method is called **linebreeding**. Linebreeding is mating individuals related to a common ancestor to maintain a substantial relationship with this animal while keeping inbreeding levels at an "acceptable" rate. These matings are usually called "elite matings" and are performed by animal breeders within their herds; on the other hand, breedings within dairy farmers' herds are called "normal matings" because their sole purpose is to give birth to milk cows and are usually performed using sexed semen[3].

The relationship between both parties is not rigid; there are cases of upward mobility between genetics companies and breeders because of proving cows as soon as they begin lactating (shortly after their first calving). Similarly, breeders are also open to incorporating animals from other parts of the world to diversify their gene pool; those animals are usually from other countries such as

---

[2]For example, the connection between polledness (presence of horns) and inheritance was established as early as 1905 (Lush, 1956), with many more traits following suit in the following years.

[3]Sexed semen contains an enriched proportion of X chromosome-bearing sperm cells to maximize the likelihood of giving birth a female animal.

New Zealand, Canada, Italy, or the UK.

Animal breeders and genetics companies compete to offer an animal with the highest trait values, which provides the best possible improvement over the cohort's average. But, as breeders continue to expand the supply of more demanded lines, the average degree of relatedness of all animals in the market grows, which leads to detrimental effects in the long run. On the other side of the market, dairy farmers want highly productive cows, so they are willing to pay a premium for the genetics from the "best" bulls.

2.2 Modeling animal breeders

An animal breeder can be modeled as a firm supplying a product ("animal genetics"), an inseparable bundle of traits embodied in a single animal. To produce such an animal, the breeder combines a set of inputs that can be split into two groups: a series of inputs exogenously supplied by the producer (labor, feed, capital, veterinary services) and a series of inputs embodied in the animal (traits.)

Melton et al. (1994, 1979) describe a breeder's decision problem as a profit-maximizing firm that combines exogenous inputs ($x_i, i = 1, \ldots, m$) and genetic inputs ($g_j, j = 1, \ldots, n$) to produce a bull y in a production function $f(.)$ to maximize profits for given bull prices (p), and input costs (w_i and r_j), then the profit function is:

$$\pi^{breeder} = pf(x_1, \ldots, x_m, g_1, \ldots, g_n) - \sum_{i=1}^{m} w_i x_i - \sum_{j=1}^{n} r_j g_j + \sum_{j=1}^{n} \lambda_j(\mu_j - g_j) \tag{1}$$

The last part reflects that breeders profit from supplying animals that perform better than the average trait level for the breed in a given period. Then we define μ_j as the average of trait j across all animals (N) in the market at a given period:

$$\mu_j = \sum_{i=1}^{N} g_{ij} f(g_{ij})$$

$f(g_j)$ is the distribution of trait j. A breeder will be paid a premium if his animal offers a degree of improvement over the average for the trait, and this is evident from the first-order conditions:

$$\frac{\partial \pi}{\partial x_i} = p\frac{\partial f}{\partial x_i} - w_i = 0 \quad i = 1, \ldots, m$$

$$\frac{\partial \pi}{\partial g_j} = p\frac{\partial f}{\partial g_j} - r_j - \lambda_j = 0 \quad j = 1, \ldots, n$$

$$\frac{\partial \pi}{\partial \lambda_j} = \mu_j - g_j = 0 \quad j = 1, \ldots, n$$

The sign of the lagrangian constraint depends on the difference between the breed average for trait j and the trait value for the bull g_j. If a breeder cannot supply an improvement over the breed average, then the marginal cost of trait j will be higher:

$$p\frac{\partial f}{\partial g_j} = r_j + \lambda_j \tag{2}$$

Achieving a marginal change in the mean level of a trait j is worth $-\lambda_j$ in terms of additional profits. Similarly, we can interpret λ_j as the Karush-Kuhn-Tucker condition that $\lambda_j > 0$ if and only if $\mu_j - g_j > 0$ and $\lambda_j = 0$ otherwise, that is, this parameter is the shadow price of genetic gains[4]. Breeders compete with each other to maximize profits; to do so, they will attempt to release into the market a bull with the highest possible trait values g_j.

Typically, breeders do not select animals based on a single trait value but as a weighted average of those traits, using what is known as a **selection index** (Hazel, 1943), where the magnitude of each weight is a measure of the importance of the trait in the context of a breeding program. Throughout this paper, we argue that most breeders have one objective: to maximize the value of an index because the higher its value, the higher the price of an animal will be.

In other words, let I be a selection index with weights ω_j, then its value will be:

$$I(\mathbf{g}) = \sum_{j=1}^{k} \omega_j g_j \tag{3}$$

With $\sum_j \omega_j = 1$ and $k \leq n$. A selection index summarizes the information from a set of k genetic traits, and it is used to rank bulls so that top-ranked animals sell their genetics at a higher price. Breeders are suppliers of genetic improvement to dairy farmers.

---

[4]As Melton et al. (1994) point out, $\frac{\partial \pi}{\partial \mu_j} = \lambda_j$ is the effect on profits of a marginal increase in the population mean of trait j

2.3 Modeling dairy farmers

Dairy farmers demand genetics to improve their productivity in the next period by choosing the bulls with the best characteristics. However, the traits that constitute an animal are marketed as a "bundle" of traits embedded in each animal. Keeping our notation from the last section, let $g_j, j = 1, \ldots, n$ be the set of genetic traits, define $G(g_1, g_2, \ldots, g_n)$ be the set of traits that defines a bull (Kerr, 1984), then the farmer's production function is:

$$y = f(x_1, \ldots, x_m; G(\mathbf{g}))$$

Where x_i is the level of non-genetic input i and G is the genetic value of the bull chosen based on their genetic traits and the animal's price, then the farmer's profit function is:

$$\pi^{farmer} = p_m f(x_1, \ldots, x_m; G(\mathbf{g})) - \sum_{i=1}^{n} w_i x_i - pG(\mathbf{g})$$

p_m is the price of their output (gallons of milk, for example), p is the price of bull genetics paid by the farmer, and w_i is the price of non-genetic input i. The most salient characteristic of this profit function is that $G(\mathbf{g})$ is discrete since there are only a limited number of options available, then:

$$\frac{d\pi^{farmer}}{dG} = p_m \frac{\partial f}{\partial G} - p = \sigma$$

The economic value of an animal's genome can be obtained by totally differentiating π^{farmer} with respect to the inputs and $G(.)$. Still, this term is not equal to zero, so $p + \sigma$ is the maximum willingness to pay for a bull's genetics, and σ is the opportunity cost of buying a specific animal's genetics. In our case, the value of σ is a decreasing function of the rank of an animal; the higher ranked an animal is, the lower the opportunity cost of his genetics.

$$p_m \frac{\partial f}{\partial G} = p + \sigma$$

$G(\mathbf{g})$ is an index of embodied characteristics, and it is assumed to be linearly homogeneous, hence:

$$pG(\mathbf{g}) = \sum_{j=1}^{n} p \frac{\partial G}{\partial g_j} g_j = \sum_{j=1}^{n} p_j g_j$$

Where p_j is the implicit price of trait j, this is the main result from the **Input Characteristics Model** of Ladd and Martin (1976), and each price is equal to the market price a trait would have if it could be marketed independently.

2.4   Genetic improvement

We have overlooked one key element of genetic change: its dynamics. We claim the animal breeders are paid according to their animal's contribution to the genetic improvement of dairy farmers' herds and the expected change in genetic traits across one generation. This rate of genetic change is affected by several factors, specifically: heritablity (h^2), selection intensity (ι), and generation interval (L) (Kerr, 1984).

$$\Delta g_{t+1} = \frac{h^2 \times \iota \times \sigma_y}{L} \tag{4}$$

Heritability is a population parameter specific to each trait (Bourdon, 2000, p. 165), and so is the standard deviation of the observed trait values. **Selection intensity** is the differential in average trait values between selected animals and the general population at time t in standard deviation units of y (observed values).

$$\iota = \frac{\overline{g}_t^s - \mu}{\sigma_y}$$

Selection intensity measures how "choosy" breeders are in deciding which individuals are selected; more intensity of selection means that only the "best" animals (according to some criterion) are selected to have offspring. If animals were randomly selected, there would be no difference between $\overline{g}^s$ and μ. Replacing these values into equation 4:

$$\Delta g_{t+1} = \frac{h^2}{L} \left(\overline{g}_t^s - \mu \right) \tag{5}$$

Δg_{t+1} is the genetic progress in a single generation of selection. Selection intensity and genetic variation are inversely proportional to the generation interval; the generation interval is the time required to replace one generation with the next; the shorter the interval, the faster the genetic change. In other words, genetic improvement is driven by two forces: selection intensity and generation interval. An increase in selectivity or a decrease in the generation interval increases genetic gains per generation.

2.5 Inbreeding Depression

A final element in our analysis is the main consequence of increased selection intensity: the mating of close relatives. **Inbreeding** is the mating of individuals more closely related to each other than the average for the population(Bourdon, 2000, p. 337). In the case of dairy cattle, the objective of mating a bull with a cow that is, to a degree, related is to preserve some of the characteristics of the bull or to help create a certain distance to a particular ancestor.

Inbreeding is not deleterious *per se* but has two consequences that may be negative. First, it increases the likelihood that an animal's offspring will perform similarly to its ancestors; for example, if a particularly tall bull is mated with a related cow, its offspring will likely be taller than average because the genes that carry such trait are likely to be present in both parents. Second, it favors the expression of deleterious recessive alleles (Charlesworth and Willis, 2009). Since an animal carries two copies of a recessive gene from each parent, it is more likely to express any such traits than an outbred animal.

In the case of polygenic traits (traits that are a consequence of the interaction among several genes), these effects are individually minimal but, in the aggregate, can have a significant negative impact on the expression of a trait; this is known as **inbreeding depression**, the result of a poor gene combination value.

Inbreeding depression is a consequence of inbreeding, but the reverse does not hold; mating closely related animals is a common way to ensure that the next generation will have similar traits to their parents. Furthermore, Falconer and Mackay (1996) show that the average level of a trait and the level of inbreeding is not linear if there are interactions between loci (positions on a specific chromosome where a gene is located.) Consequently, an increase in the level of inbreeding does not necessarily imply a higher likelihood of inbreeding depression, but as the population becomes more inbred, it will be the case.

Inbreeding is commonly measured using the **Inbreeding Coefficient** (F_X) (Wright, 1922), defined as the probability that two alleles[5] in an individual were both descended from a single allele in an ancestor.

$$F_X = \sum_{ca=1}^{K} \left(\frac{1}{2}\right)^{n_1+n_2+1} (1 + F_{ca}) \tag{6}$$

Where ca denotes the common ancestor, n_j is the number of generations separating the com-

---

[5]Alternative forms of a gene, inherited from the sire, the dam or both.

mon ancestor from the sire (dam), and F_{ca} is the inbreeding of the common ancestor.

3   Overview

3.1   The North American cattle genetics market

The United States and Canada together make up the largest cattle genetics market in the World; about 2/3 of total bovine semen exports originated in one of these markets in 2019.
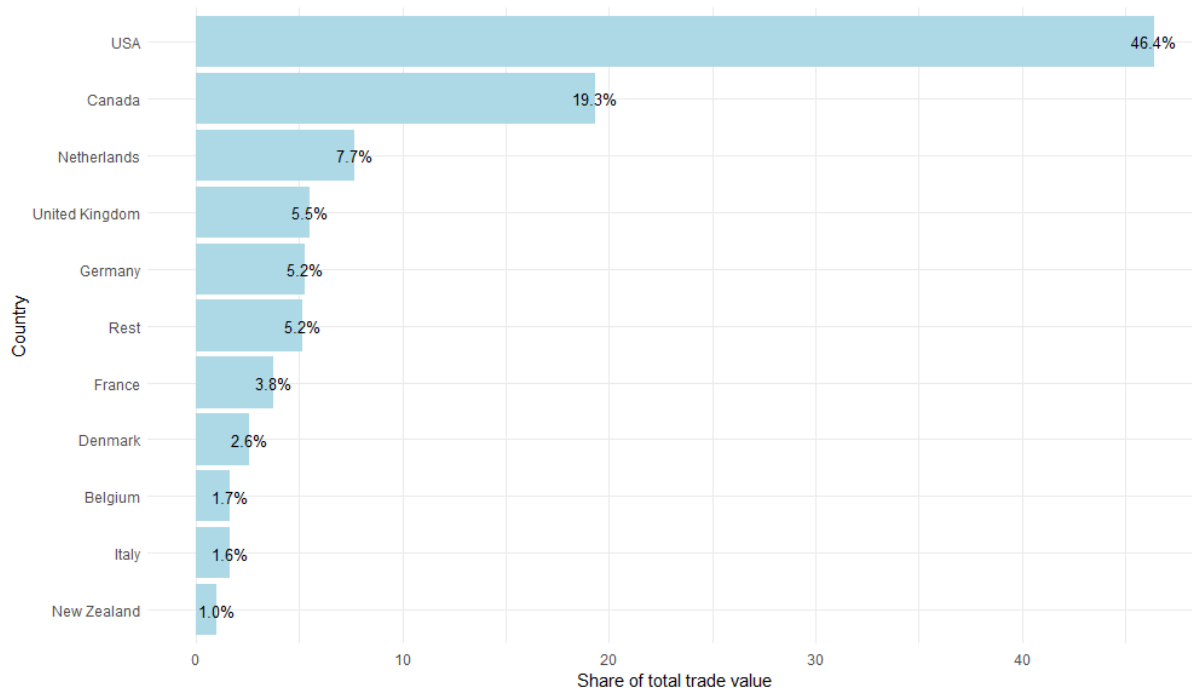

Figure 1: Share of trade Value of Bovine Semen exports by country (2019)
Source: UN Comtrade database

As a consequence, any technological innovation that may have an impact on this market will have consequences throughout the world. For example, the development of artificial insemination in 1938 and freezing technology for bull semen in 1953 brought into existence a group of animals we call "Superstar Sires", bulls who have fathered thousands of daughters because of their superior genetics and gave birth to the three dominant pedigrees of the present day, namely those of "Pawnee Farm Arlinda Chief," "Round Oak Rag Apple Elevation," and "Penstate Ivanhoe Star." (Hutchins and Hueth, 2023)

Two (exogenous) technological innovations in the 20th century led to a significant decrease in the male diversity of Holstein cattle[6]. According to Yue et al. (2015), all Holstein bulls in North

---

[6]The Holstein or Holstein-Friesian breed, originated in the present-day Netherlands and it is the most widespread

America are descendants of two bulls imported from the Netherlands in 1880: "Neptune H" and "Hulleman."

This article will investigate the consequences of a different technological innovation: Genomic Selection for bulls. Since the entire genome of the *Bos Taurus* was sequenced in 2007, genetics companies can identify which bulls have the highest value among a set of quantitative traits and can pass them to their offspring.

3.2   Genomic Selection

Selective breeding is the only way to carry out directional genetic change (Melton et al., 1979); by doing so, breeders decide which bulls mate with a specific cow to produce the next generation of animals whose semen will be marketed to dairy farmers, who are interested only in cows, since every all production traits (e.g., milk and protein content) are expressed in females only.

Traditional selection techniques relied on information from pedigrees and observed trait values to assess an animal's fitness. Genomic selection uses data from an animal's genome to accurately predict breeding values, particularly in cases where traits are expressed only in females. Indicating the value of an animal as a parent (Breeding Values) in the case of traits that are expressed only in females used to require evaluating a large number of daughters from the same bull; for that to happen, these cows must have reached the age of sexual maturity (one year); hence, the entire process could take three to four years until enough information was gathered.

Instead, genomic selection uses the bull's DNA to detect gene combinations significantly different than the expected frequency if they were associated at random (Goddard and Hayes, 2007) to detect the presence of certain genetic traits that are highly correlated with specific gene combinations. This process can be undertaken as soon as the calf is born, so by the time the animal has reached sexual maturity (one year old), it can be marketed immediately.

In 2009, the whole bovine genome was published (Elsik et al., 2009), providing animal breeders with the last missing piece required to make genomic selection commercially viable; the "50K" chip was released that same year, a commercial system that simultaneously detects up to 50,000 SNPs, thus kickstarting the genomic selection era. Genomic testing of young bulls immediately decreased the generation interval (average age of parents when their select offspring is born). It first accelerated the genetic gains (variation in breeding values per year) for North America's dairy cattle population and then the rest of the world. Larger genetics companies adopted ge-

---

dairy cattle breed in North America since it was introduced in 1852.

nomic selection techniques faster because of their sunk costs.

We hypothesize that increased competition for genetic improvement implies more competition for "superstar" lines, that is, bulls descended from a notorious ancestor (e.g., bulls that fathered many daughters). The rationale of these choices is that the more related two animals are, the more likely their offspring is to preserve their genetic traits. Higher inbreeding levels lead to a higher likelihood of observing **inbreeding depression**, a decrease in the performance of inbred animals due to lower fertility, higher incidence of rare diseases, and reduced longevity.

The link between genetic improvement and inbreeding rates can be shown by looking at the **Total Performance Index** (TPI), a selection index used by the Holstein Association of the United States of America. Figure 2 shows a change in the trend starting in 2010 but leveling after 2018.
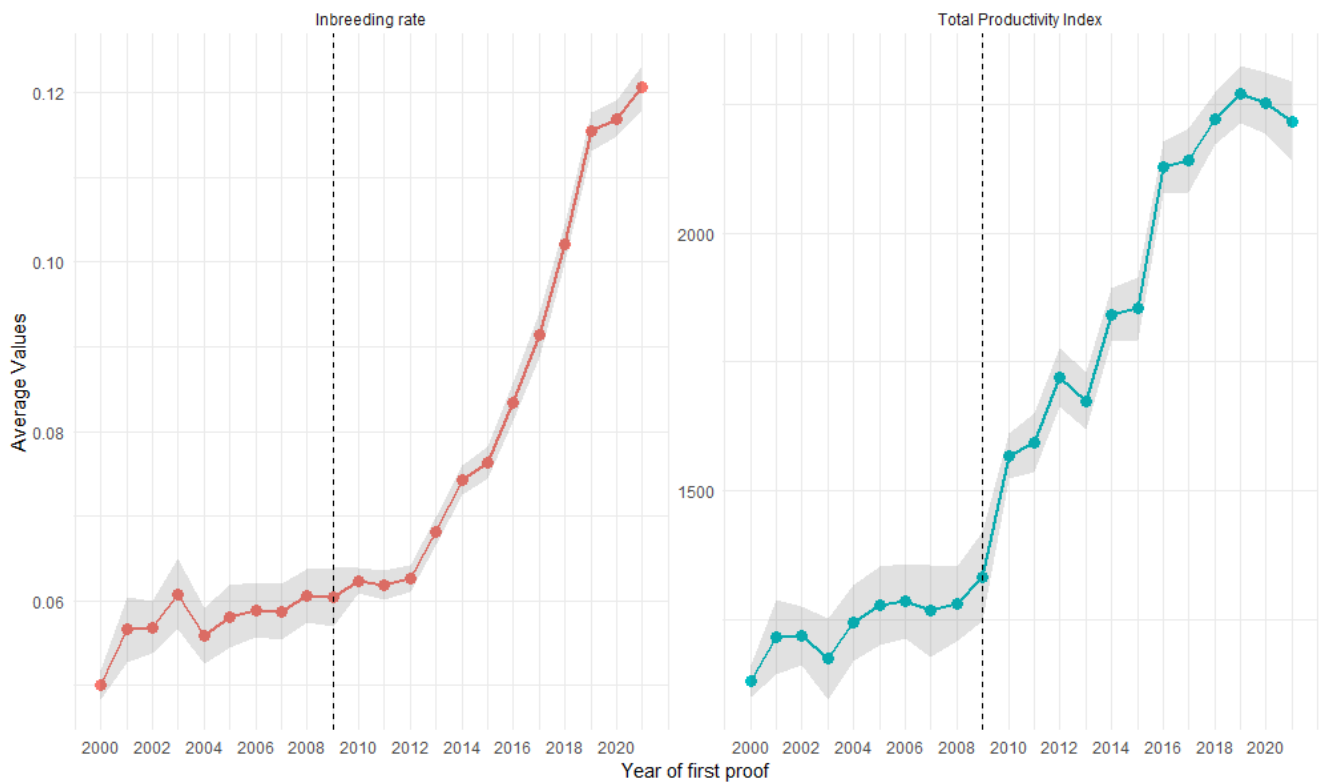

Figure 2: Average Total Performance Index of Holstein bulls
Source: NAAB

With genomic testing, it is no longer necessary to wait for the bull's progeny to grow since an animal can be tested as soon as it is born for the presence of specific traits on its genome. Consequently, the generation interval, the average age of parents when offspring are born (Wiggans and Carrillo, 2022), has decreased from around 5.5 years to less than two years, close to the biological minimum.
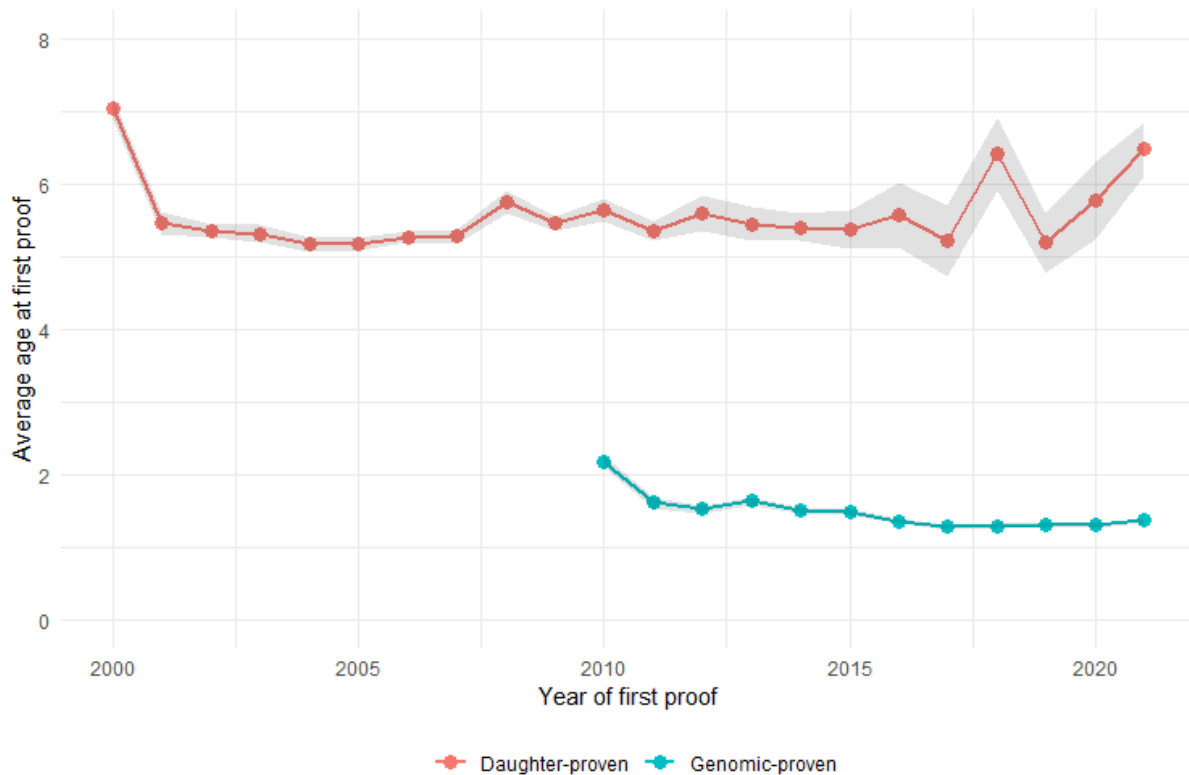
Figure 3: Average age at first proof for Holstein bulls
Source: NAAB

The most important consequence of this shorter generation interval is the increase in genetic progress. According to García-Ruiz et al. (2016), the most significant genetic gains were observed in lowly heritable traits such as DPR (Daughter Pregnancy Rate), PL (Productive Life), and SCS (Somatic Cell Score).

Figure 4 shows the relationship between inbreeding and production traits (milk yields, percent of fat in milk, and pounds of protein in milk) split between three periods, pre-2009 (before genomic selection), 2010-2015, and 2015-2020, and by the status of the animal[7].

[7]"Active" means that the animal's traits were tested on his daughters, "Genomic" implies that they were tested through the bull's genome, "Foreign," means that the animal was tested outside the United States.
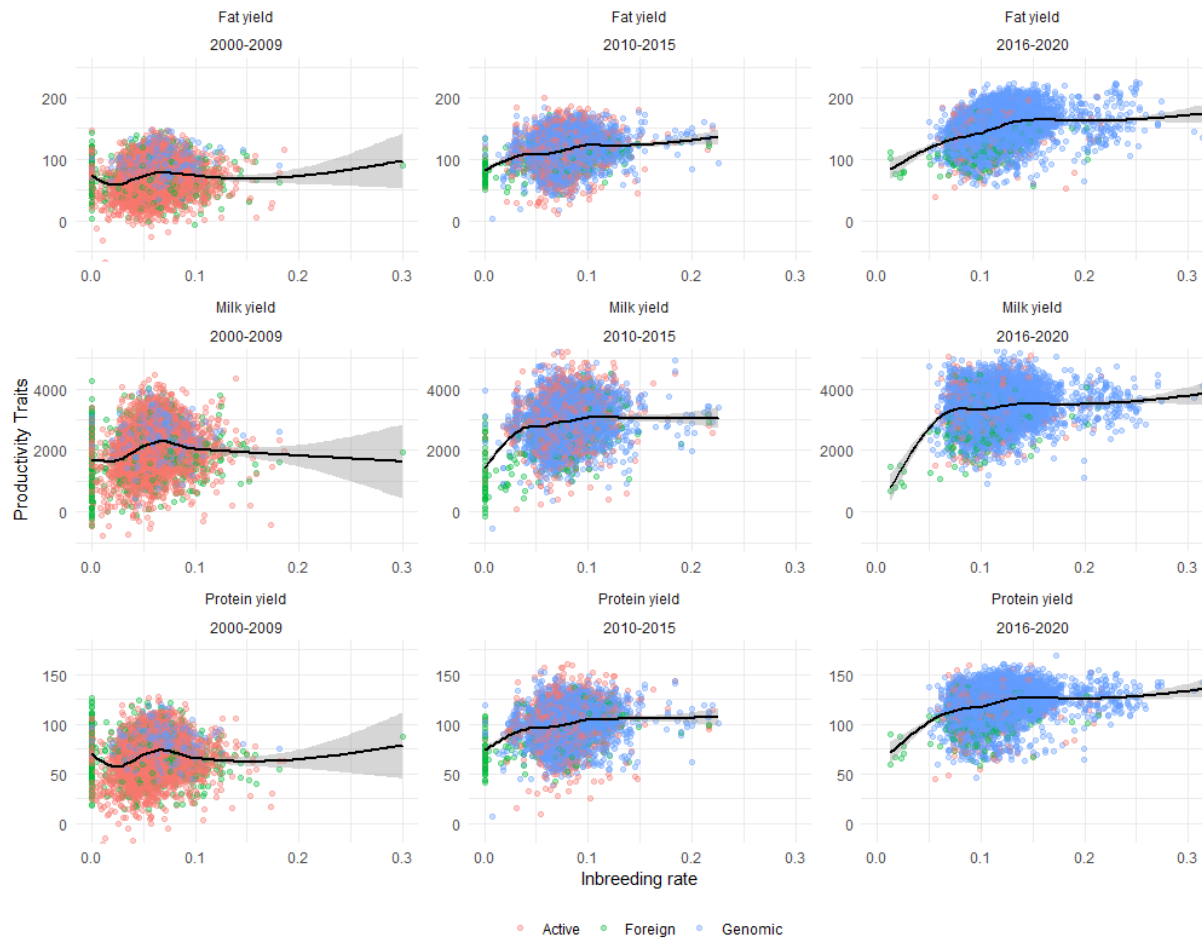
Figure 4: Production trait values for Holsteins as a function of inbreeding rate
Source: NAAB

The relationship between production trait levels increases with the inbreeding rate, as animals are partially selected for them. The strength of the relationship varies with inbreeding levels; for values between 0.05 and 0.15, there is a positive relationship between inbreeding and average values of the three traits. However, as inbreeding levels increase, the strength begins to plateau.

However, as inbreeding rates continue to increase, inbreeding depression becomes a concern (Cole, 2019), regarding fertility traits. Figure 5 shows the relationship between the inbreeding rate and three fertility traits: Daughter Pregnancy Rate (DPR), the percent of nonpregnant cows that become pregnant during 21 days; Cow Conception Rate (CCR), the percentage of inseminated cows that become pregnant each insemination; and Heifer Conception Rate (HCR), the percent of inseminated heifers that become pregnant after each insemination. The Daughter Pregnancy Rate measures how easy it is for the bull's daughters to become pregnant. In contrast, cow and heifer conception rates measure how likely a cow or a heifer will become pregnant by artificial insemination. Fertility traits are the ones more likely to be affected by inbreeding depression.

Still, the plot doesn't hint at apparent effects other than a higher average inbreeding rate across periods and genomic-proven bulls having more variability in most traits than daughter-proven bulls.
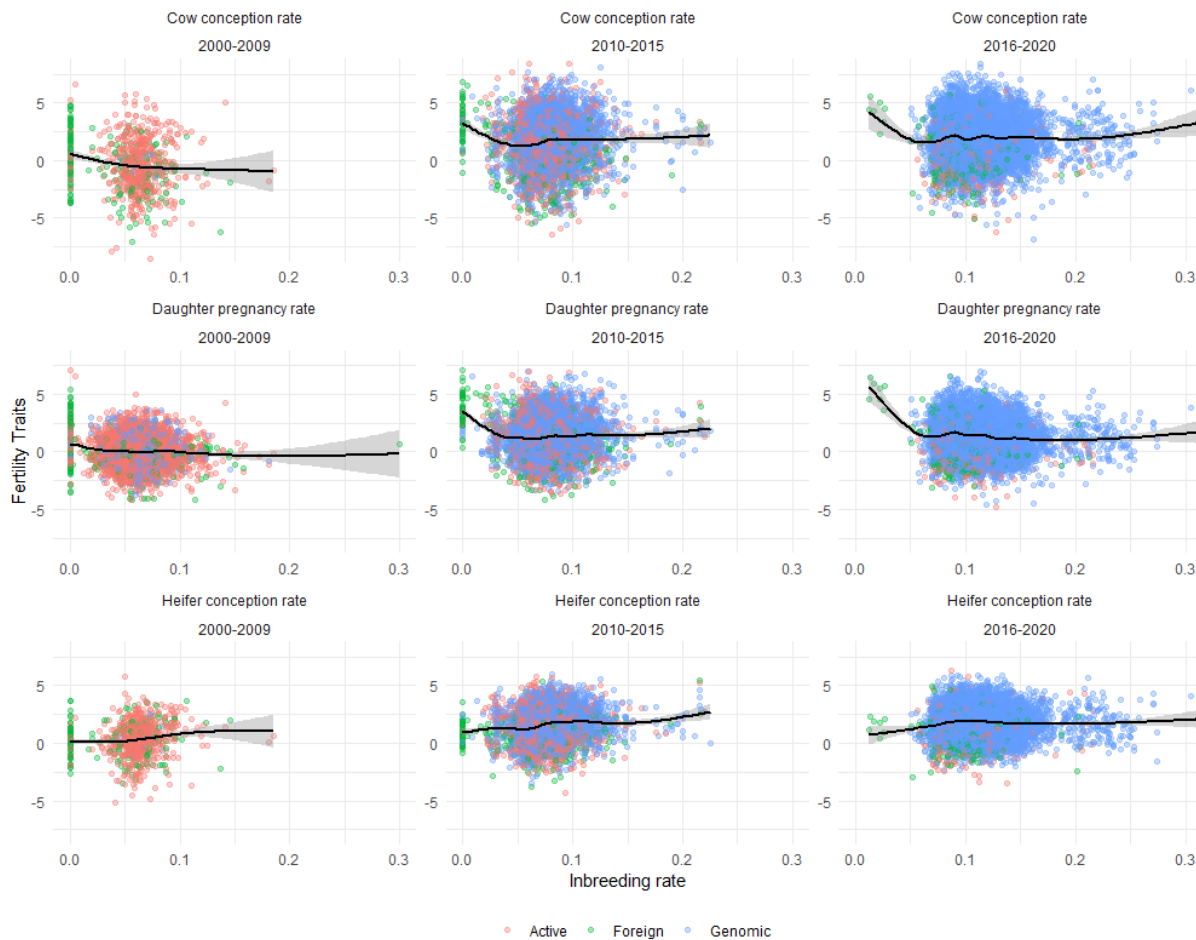

Figure 5: Fertility traits for Holsteins as a function of inbreeding rate
Source: NAAB

4 Data Description

4.1 Estimation of Genetic Traits

We now turn to a rigorous definition of a "genetic trait," that is, the portion of an observed value of an animal's characteristic that can be measured (i.e., quantitative traits, as opposed to qualitative traits such as hide color or polled status.) The observed value of an attribute is defined to be the result of two effects: a **Genetic influence**, the set of genes (genotype) inherited from the animal's parents combined (expressed) during the gestation period, and an **Environmental effect**, a term that is used to capture a series of any external (non-genetic) factors that affect the

animal, such as the food it consumes or the characteristics of the farm where it resides.

In the American Dairy industry context, the National Dairy Herd Improvement Program (ND-HIP), established in 1925, systematically collects, manages, and analyzes data from dairy cattle (Hutchins and Hueth, 2023). The information it collects, commonly called "evaluations," is used to assess the industry's productivity and estimate the paternal contribution to these traits. The raw data collected from farms is used to estimate a statistical model, called the "animal model" (Henderson, 1975, Wiggans and VanRaden, 1989), to predict the genetic contribution of an animal as a sire (or dam) after controlling for environmental variables herd, age, pedigree, and season effects. The predicted values for each trait from this model are commonly called **expected transmitting ability** (PTA) of a trait.
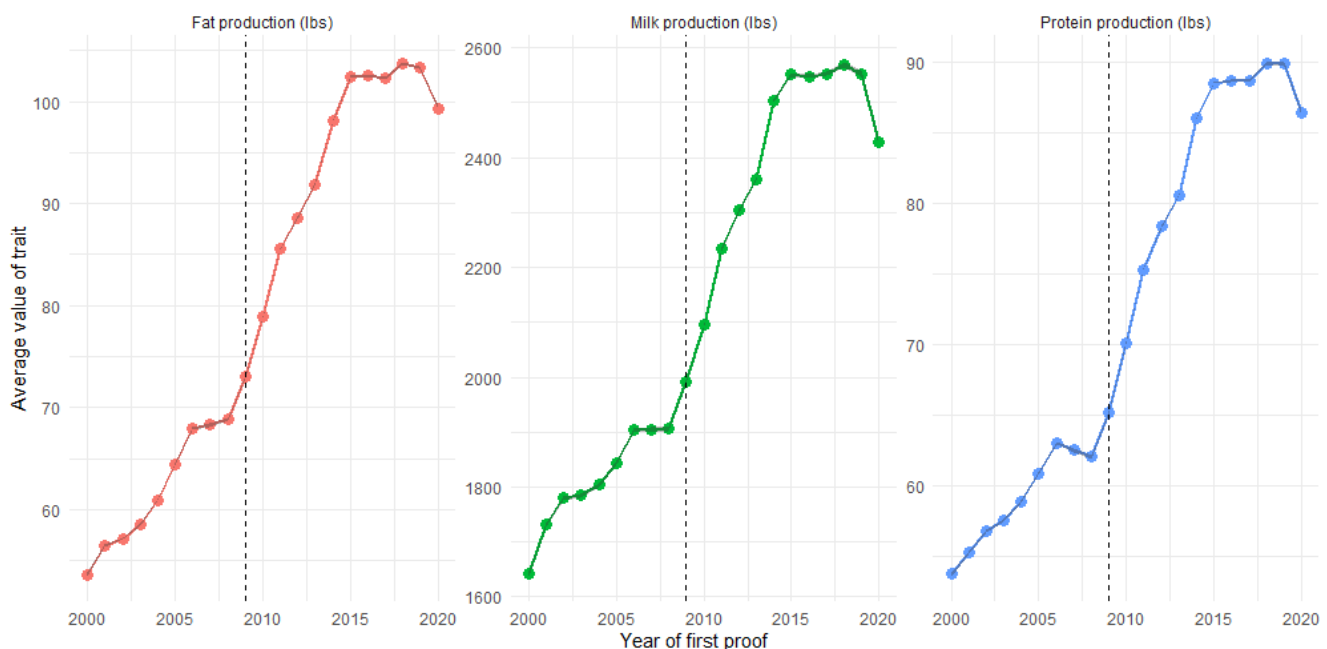

Figure 6: PTAs of production traits
Source: NAAB

4.2 Market structure and competition in the genetics market

As mentioned earlier, the supply of dairy genetics comprises animal breeders and genetics companies; in some cases, they are vertically integrated, but this is not necessarily the case. These companies sell semen to dairy farmers, who, in turn, use it to impregnate their cows.

Perhaps the closest analogy is the market for seeds; in that market, seed companies develop new varieties based on farmer's demand, such as glyphosate or drought tolerance, and then sell seeds to farmers who plant them. Seed companies then developed genetically engineered traits

subject to copyright law, thus forcing farmers to sign restrictive retailing contracts to replant those seeds in further seasons (Ciliberto et al., 2019).

The dairy genetics industry works similarly when a dairy farmer buys semen; they expect to obtain a female. However, if that is not the case, the genetics company usually has the right to buy back the calf to avoid being outbid by a competitor.

Animal breeders and genetics companies compete to offer an animal with the most significant possible improvement over the cohort's average. As a consequence, firms will try to produce bulls with the highest productivity traits; similarly, this will lead to higher rates of inbreeding since the best way to ensure high productivity is through linebreeding, that is, by building a pedigree that has a relatively stable degree of relationship relative to a specific ancestor.

However, the more inbred selected animals are, by definition, also increases the inbreeding of an average animal in any cohort, thus making inbreeding depression more likely within a herd. The effective population size[8] of bulls is at a historic low; according to Makanjuola et al. (2020), the effective population size of Holstein bulls ranges from 43 to 66 animals. One way to tackle this problem is by importing foreign bulls with lower trait levels, which can diversify the gene pool.

The first official genomic evaluations were released in January 2009 for Holsteins and Jerseys (Wiggans and Carrillo, 2022). However, not all genetic companies adopted the technology simultaneously. Genomic testing allows genetic companies to detect suitable bulls earlier; thus, they can release those animals into the market as soon as they reach sexual maturity. Figure 7 shows the total number of new bulls released in the market by year and type of firm.

---

[8]this is, the number of individuals that an idealized population would need to have for inbreeding rates to be the same as in the actual population.
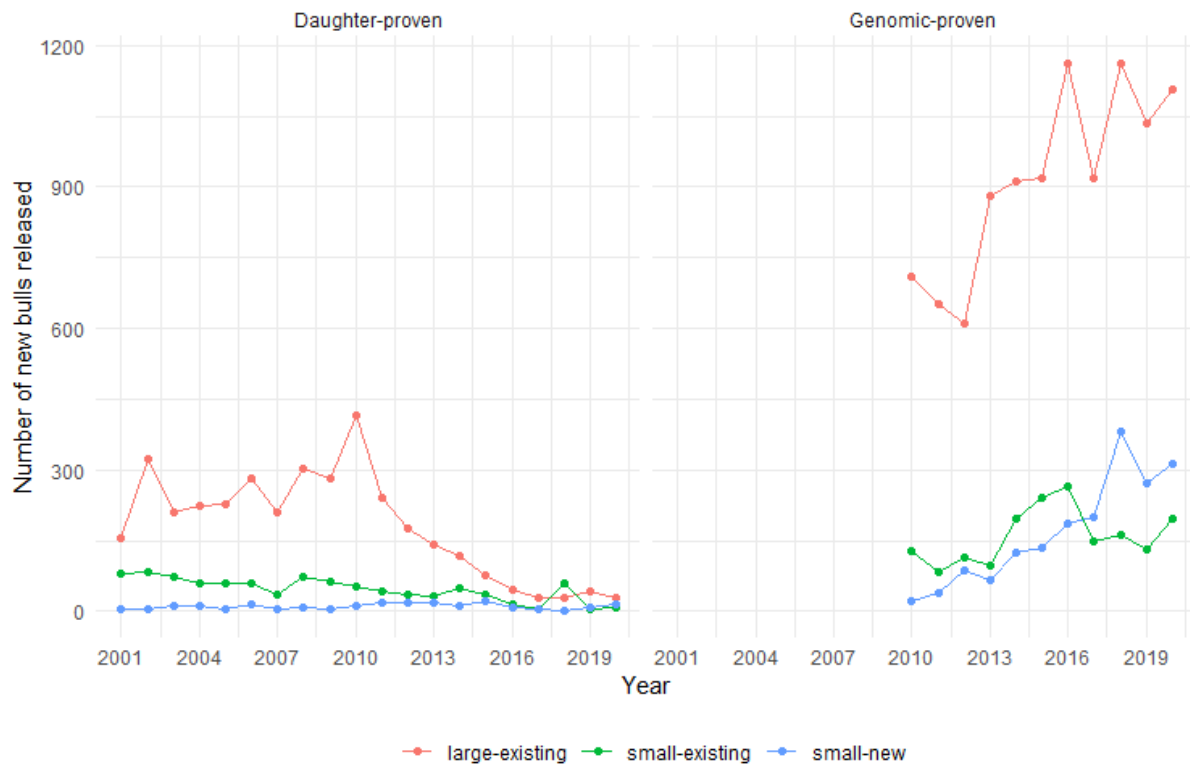
Figure 7: Number of genotyped bulls by firm size
Source: NAAB

We call "large" companies those with the largest share of the North American genetics market. Similarly, a firm is called "preexisting" if it sold bull genetics before 2009; otherwise, they are considered "new."
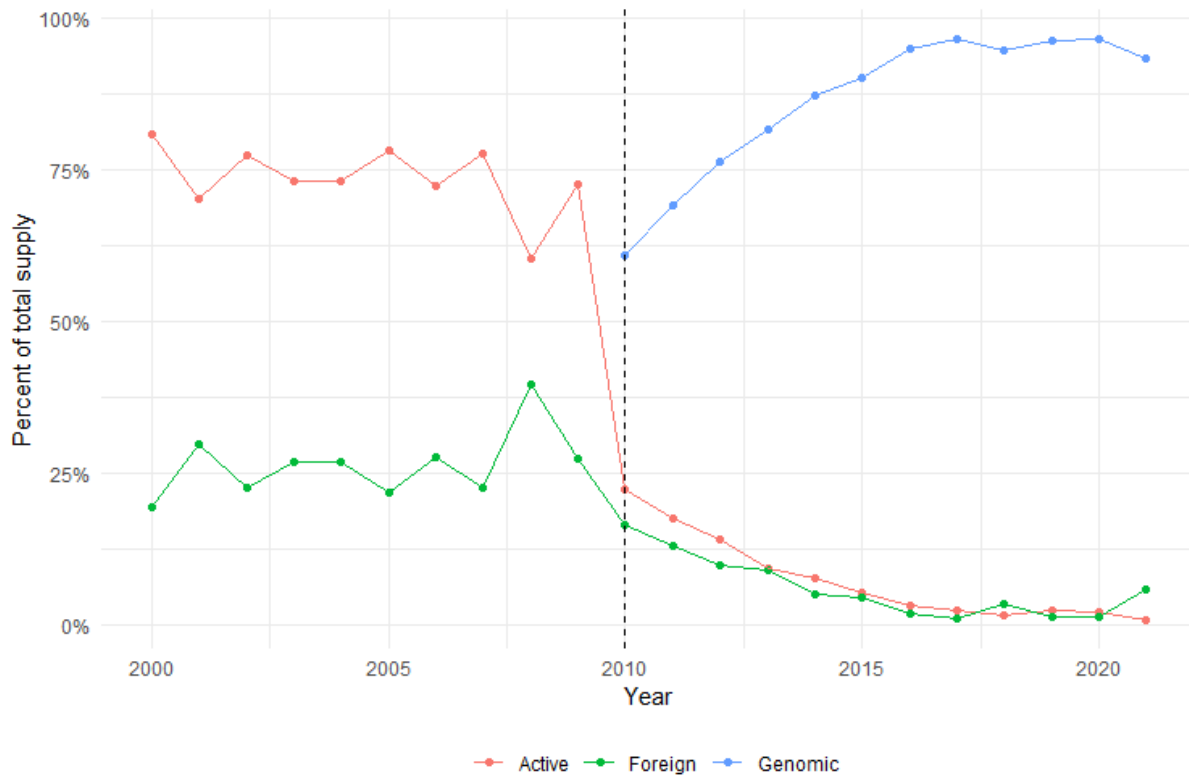
Figure 8: Share of each status of bull by year

In the cattle genetics industry, genomic testing has been adopted remarkably fast. Figure 8 shows how fast genomic testing was adopted in our sample. Technology adoption is usually modeled as a slow-moving process (Hall and Khan, 2003) resulting from aggregating several individual decisions depending on each decision maker's information, uncertainty, and cost structure. Thus, adoption patterns tend to follow an S-shaped curve (Griliches, 1957), where a relatively small number of agents adopt the new technology first, but as time progresses, more and more people adopt it. It grows exponentially until it reaches 100%.

This is not the case, and the high starting costs for genotyping can easily explain it since large companies were the ones that could initially afford it. Still, since the cost decreased quickly, more and more companies could genotype their calves. Figure 9 breaks down the adoption percent by firm size; larger firms genotyped a more significant percent of their newborn calves in 2010, and the share has grown steadily, while smaller firms followed them at a slower rate.
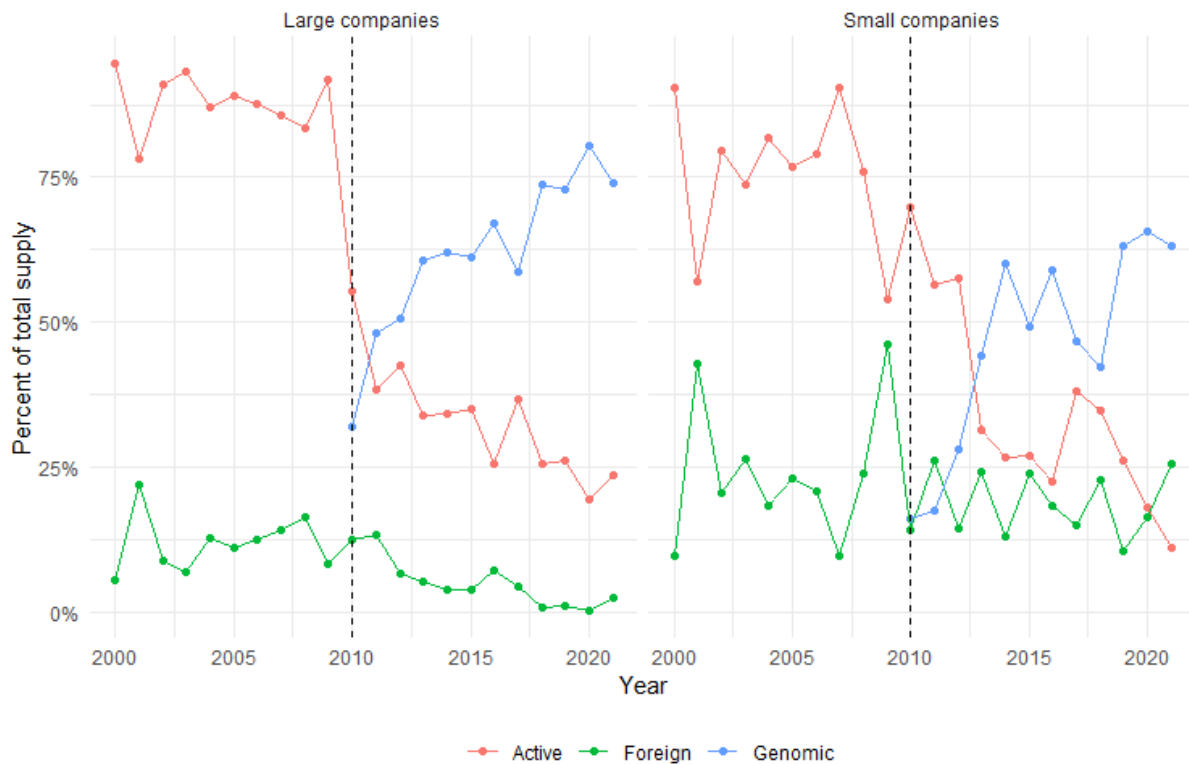
Figure 9: Share of each status of bull by year and company size

Starting in 2010, large firms started genotyping their bulls, and consequently, the number of new bulls increased threefold in that year, when all bulls aged 1 to 4 were released simultaneously. We argue that this staggered adoption of genomic selection has increased preexisting trends for consolidation in the market and has also led to more competition for "superstar" lines of bulls.

However, the more inbred selected animals are, by definition, also increases the inbreeding of an average animal in any cohort, thus making inbreeding depression more likely within a herd. The effective population size[9] of bulls is at a historic low; according to Makanjuola et al. (2020), the effective population size of Holstein bulls ranges from 43 to 66 animals. One way to tackle this problem is by importing foreign bulls with lower trait levels, which can diversify the gene pool.

The first official genomic evaluations were released in January 2009 for Holsteins and Jerseys (Wiggans and Carrillo, 2022). However, not all genetic companies adopted the technology simultaneously; larger firms adopted genetic selection earlier. As mentioned, genomic testing allows genetic companies to detect suitable bulls earlier; thus, they can release those animals into the market as soon as they reach sexual maturity. Figure 7 shows the total number of new bulls re-

---

[9]this is, the number of individuals that an idealized population would need to have for inbreeding rates to be the same as in the actual population.

leased in the market by year and type of firm. We call "large" companies those with the largest share of the North American genetics market. Similarly, a firm is called "preexisting" if it sold bull genetics before 2009; otherwise, they are considered "new."

5   Empirical Framework

5.1   Lines

We define a **line** as the set of direct (male) descendants from a "superstar" sire, which, in turn, can be defined in terms of his number of male or female descendants. We use two definitions of superstars; in the first one, a bull belongs to this group if the size of his progeny is in the 95th percentile of the distribution of female descendants. In the second case, a bull is a superstar if his number of sons is in the 95th percentile of the distribution of male descendants.

We argue that genomic testing of bulls made breeders "double down" on existing lines rather than investing in crossbreeding or new, less inbred lines. Daetwyler et al. (2007) argued to the contrary, that genomic selection would reduce inbreeding since it allows for a more accurate estimation of breeding values, and this greater accuracy would allow breeders to 'discover" new potentially high-performing sires independent from their ancestry.

Breeders are under high competitive pressure to supply bulls with the highest PTAs. However, the best way to do so is by linebreeding a family of bulls so that the relative distance to an ancestor remains constant across generations. Steyn et al. (2023) shows that younger sires capture a more significant share of present-day population genetic variation; hence genotypes of older sires can be discarded from (genomic) evaluations. As generations begin to accumulate, the contribution of a particular sire decreases (on average) exponentially, so we can disregard the impact of distant ancestors.

We include in the sample all animals whose sires were born between 1996 and 2005, and we call those sires the "founders" of each line. Our definition depends critically on the period when the founders of such lines were born; if we choose an early date, we make most animals in the sample related to each other to a degree; if we choose a later date instead, our lines will be too short to make any statistical inferences. Every year, a new generation of animals is born from any line, each with a given inbreeding rate and a set of genetic traits inherited from his parents, and we can estimate the mean and standard deviation of inbreeding rates per line.

5.2 Matching "supersires"

After identifying the sires, we can recursively build their pedigrees using their registration numbers until we find the latest animal in the sample. We want to measure the causal effect of the introduction of genomic testing in 2010 on the average inbreeding in extant lines; the first problem we must tackle is the nature of a valid comparison group. We argue that the introduction of genomic testing caused animal breeders to supply more animals from well-known lines, that is, descended from superstar sires; our control group must be the lines descended from other non-superstar sires.

First, we need to match the founders of such lines with other sires that did not father as many sons or daughters but had comparable genetic traits; then, we find all descendants from both groups and compare the evolution of their inbreeding rates through time. The idea behind this procedure is that when those bulls were in the market, any reasonably similar sire could have been chosen but was not, and thus can be used to construct a counterfactual line.

For every family, we must construct a counterfactual family with new bulls being born yearly, each with an inbreeding rate that we can average across bulls from every cohort.

Then, we define the **treatment group** as all lines (families) of animals that are comprised of direct descendants of "superstar" sires. Similarly, our **control group** comprises descendants from animals matched to superstar bulls regarding their genetic traits.

Our next objective is to construct counterfactual (synthetic) lines by matching superstar sires with others that were not as successful but had similar genetic traits. To do so, we use a Propensity Score Matching Estimator, let $p(X) = \mathrm{Pr}(superstar_i = 1/X = x)$ where $superstar$ is a dummy equal to 1 if bull i's number of daughters is on the 95th percentile of the progeny distribution and X_i be a vector comprised of the following:

1. Predicted Transmitting Ability for production traits (milk, fat, and protein), health traits (mastitis, metritis, productive life), and reproductive traits (daughter pregnancy rate, effective calving, gestation length).

2. Dummies for the absence of certain recessive genes (Holstein Foundation, 2021) such as *brachyspina*, cholesterol deficiency, and mule foot.

3. Dummies for the presence of genes (polled, red coat color, complex vertebral malformation).

4. Dummies for the presence of haplotypes (specific sequences of DNA at different locations

in the chromosome) that affect fertility (Holstein Foundation, 2018) when present in homozygous form (likely to be expressed when two carrier animals are mated).

We then estimate the propensity score (Cameron and Trivedi, 2005, p. 873), interpreted as the probability of being a superstar given the level of PTAs and the presence of the entire set of variables listed above.
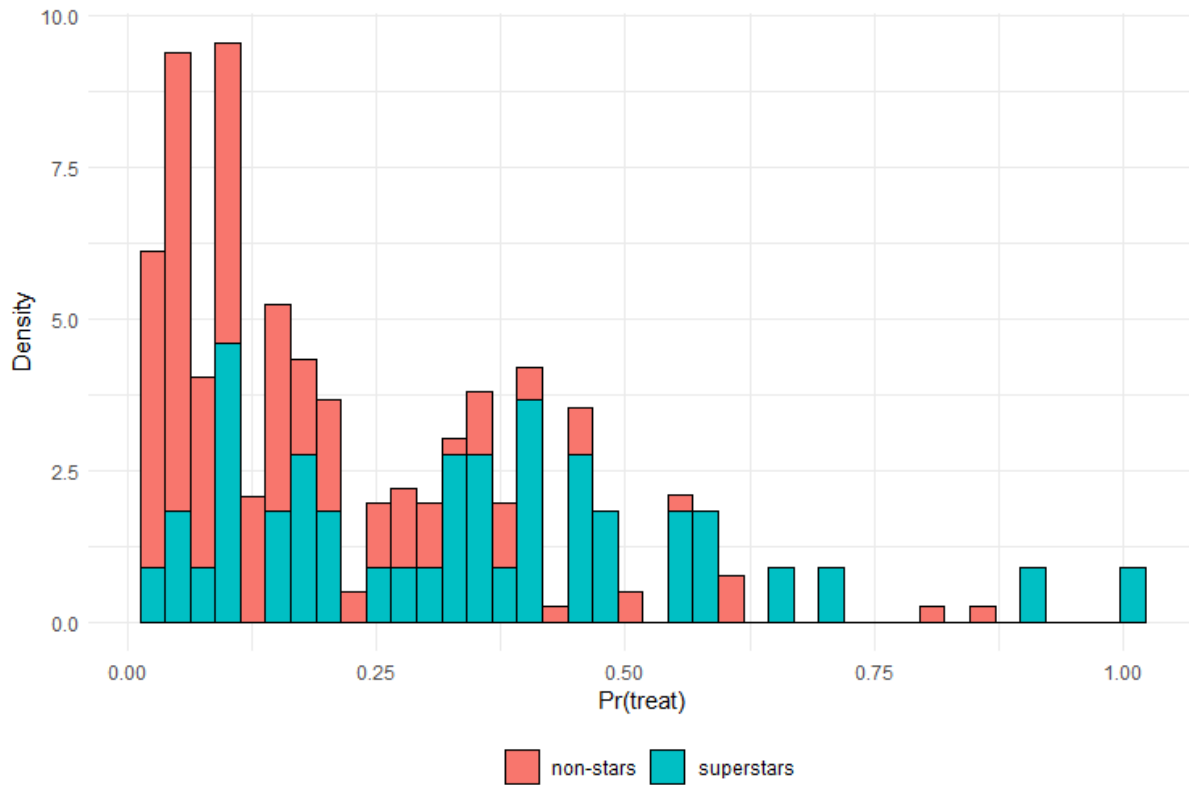

Figure 10: Propensity score by class of bull

Figure 10 shows the degree of overlapping between superstars and the remaining bulls; since there is a considerable difference in the size of the groups, superstars make up only a tiny fraction of bulls; however, the overlapping is good in the range $0.05 - 0.50$, but there exist a few "superstar" that had little close matches in the sample.

With these estimated probabilities, we use Nearest Neighbour Matching (Smith and Todd, 2005) to match superstars with twenty alternative bulls, to account for the fact that non-superstar bulls will, by definition, have significantly less progeny. Consequently, given the total number of animals and the trait levels, we want to match a superstar with as many comparable bulls as possible.

Each year these bulls will have a new set of descendants being born, each with its inbreeding rate and genetic traits; if our hypothesis holds, then the inbreeding rates of the descendants of

"supersires" should be significantly higher than those of non-superstar descendants.

6   Results

Figure 11 shows the average inbreeding rates and confidence intervals for both treatment and control groups across years. We have that genomic selection did not increase inbreeding rates immediately; both curves began to diverge substantially after 2011. Inbreeding rates did not take off immediately after the introduction of genomic testing, but after two years, both curves diverged substantially from 2012 onwards.

This plot closely matches the evolution of the population average inbreeding rates plotted in Figure 4, where the increase in inbreeding has a two-year lag to productivity growth. This behavior is consistent with our hypothesis that breeders prefer well-known over new lines. Our findings coincide with results reported by Guinan et al. (2023), who also look at the time trends of productivity and reproduction traits in US dairy cattle. The authors also document a decrease in the generation interval and an increase in inbreeding rates of Holstein and Jersey cattle but do not consider the latter to be an issue because genetic trait values are adjusted by inbreeding of future progeny.

The Holstein Association of the United States[10] uses a simple formula to adjust PTAs of several traits by inbreeding rates:

$$\overline{PTA}_i^k = PTA_i^k + (\alpha^k * F_i) \tag{7}$$

$\overline{PTA}_i^k$ and PTA_i^k are the corrected and uncorrected predicted transmitting ability of trait k for animal i, α_k is a trait-specific constant and F_i is the inbreeding rate of animal i. α_k is negative to reflect the negative impact of inbreeding on fertility.

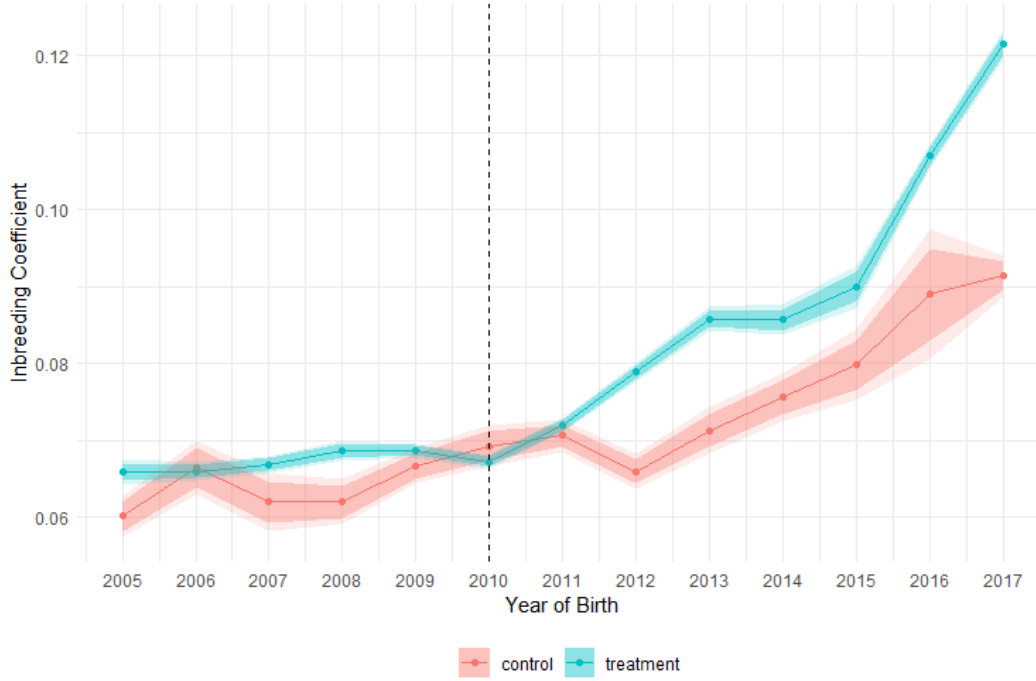[10]https://www.holsteinusa.com/pedigree_info/calc_details.html

Figure 11: Average inbreeding rates in treatment and control groups

To measure the effects of genomic selection on inbreeding, we estimate the differences-in-differences model:

$$F_{it} = \beta_1 treat_i * yob_t + yob_t + \mathbf{X}_{it}\gamma + \psi_{it} + \varepsilon_{it} \qquad (8)$$

Where F_{it} is the inbreeding rate of bull i born in year t, $\mathbf{X}_{it}$ is the set of genetic traits, yob_t is a year of birth fixed effect, and ψ_{it} is a firm fixed effect.

Table 1 shows the results of three models, model 1 is the analog to Figure 11, including only the interactions between treatment and year of birth. Model 2 includes the genetic traits, and Model 3 is the complete model with firm fixed effects.

Table 1: Synthetic difference-in-Differences estimates

	Inbreeding rate (%)			
	No covariates	No covariates	Traits	Traits and firm FEs
	(1)	(2)	(3)	(4)
treated	0.348** (0.136)	0.197 (0.247)	0.135 (0.265)	0.179 (0.263)
treated × post	0.400 (0.285)			
treated_2005		0.347 (0.408)	0.285 (0.446)	0.050 (0.459)
treated_2006		−0.275 (0.359)	−0.115 (0.384)	−0.266 (0.373)
treated_2007		0.293 (0.338)	0.323 (0.299)	0.317 (0.304)
treated_2008		0.466 (0.308)	0.354 (0.349)	0.271 (0.352)
treated_2010		−0.406 (0.312)	−0.435 (0.347)	−0.468 (0.352)
treated_2011		−0.054 (0.268)	−0.110 (0.278)	−0.123 (0.280)
treated_2012		0.933*** (0.352)	0.988** (0.404)	0.949** (0.411)
treated_2013		1.225*** (0.382)	1.003** (0.449)	0.936** (0.469)
treated_2014		1.022*** (0.339)	0.535 (0.388)	0.454 (0.383)
treated_2015		1.467*** (0.371)	1.080** (0.443)	1.053** (0.511)
treated_2016		1.597*** (0.429)	1.046*** (0.352)	0.995*** (0.326)
treated_2017		2.810*** (0.395)	2.732*** (0.344)	2.923*** (0.337)
Observations	15,627	15,627	12,153	12,153
R^2	0.427	0.434	0.467	0.485
Adjusted R^2	0.426	0.433	0.465	0.480
Residual Std. Error	1.840 (df = 15612)	1.829 (df = 15601)	1.825 (df = 12105)	1.800 (df = 12033)

Notes: ***Significant at the 1 percent level.

                    **Significant at the 5 percent level.

                    *Significant at the 10 percent level.

All three models that incorporate year fixed effects in Table 1 show a series of significant differences in treatment effects for 2012-2017; however, including PTAs into the equation makes the 2014 fixed effect insignificant. Regression models show a positive and increasing inbreeding rate in the treatment group comprised of all descendants from prestige sires relative to the control group of descendants of comparable non-superstar sires.

The last period in the sample is 2017 instead of 2020 because our control group "collapsed" in 2018, meaning that the number of comparable descendants is too small for us to consider them a valid comparison group. This drastic reduction is a consequence of lower fertility rates in the

control group; since those sires fathered fewer sons, their sons will father even fewer grandsons, and after a few years, this group collapses.
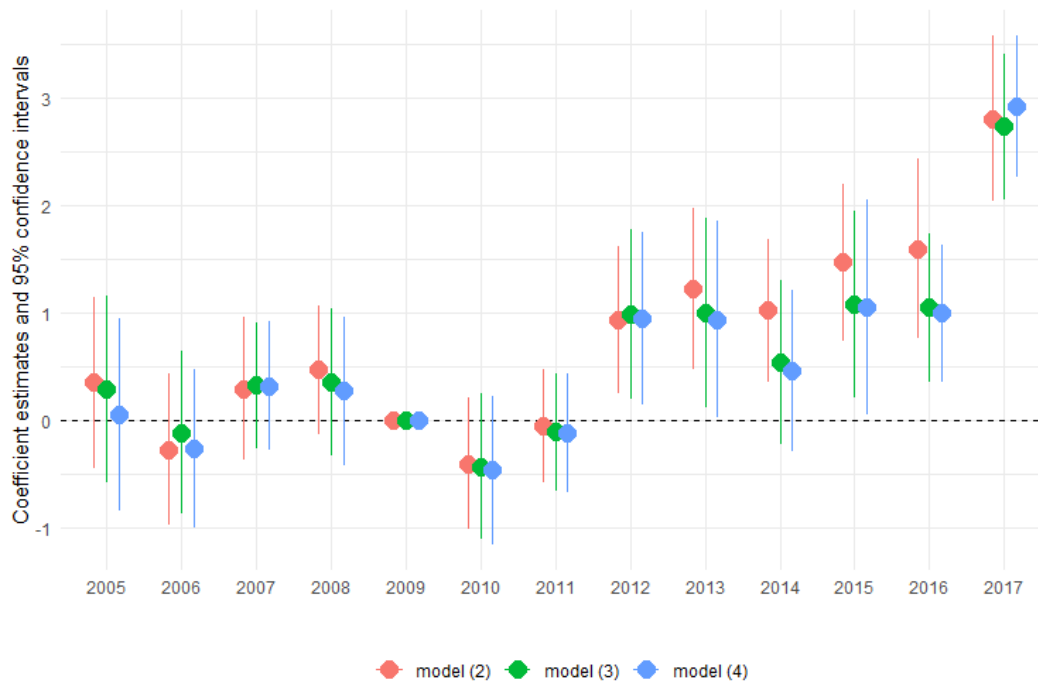

Figure 12: Coefficient plots for Regression 8

Figure 12 shows the coefficients from the interaction of year fixed effects and treatment status to assess the impact of inbreeding per year correctly; every year after 2012, the inbreeding rate of the treatment group is between 1 and 2.8 percent points higher than the control group. The coefficient values decrease slightly after controlling for their genetic traits but remain statistically significant. According to the Holstein Association data, each additional percent point of inbreeding has a $23 revenue loss per cow, implying that increased inbreeding leads to a cost per cow that ranges between $ 22 and $ 64 per animal and year. We can then multiply this value by the total number of cows in the US, 89.3 million animals[11] to get a total cost to the dairy industry that ranges from $ 1.96 to $ 5.7 billion for the entire US dairy industry.

7 Discussion

This article investigates the consequences of genomic selection on the American dairy industry. On one side, this new technology significantly improved productivity, but on the other side, it increased the average inbreeding rate of all bulls in the market.

---

[11]https://www.nass.usda.gov/Newsroom/2023/01-31-2023.php

The increase in productivity was fueled by two effects, first by a decreased generation interval motivated by a new technology that gives breeders an accurate estimation of a bull's genetic traits as soon as it is born. This faster access to information allowed breeders to select better bulls faster, and consequently, genetic gains per generation are now much higher than they used to be in the past.

Breeders are now under higher pressure to produce high-performing bulls, and the easiest known way is by breeding lines that descend from a famous ancestor; we call these animals "supersires", and we show that breeding on these well-known lines is the most critical driver of a greater degree of consanguinity among bulls. Inbreeding is not inherently bad, as the average inbreeding rate becomes higher, and every new generation of animals experiences an increased risk of experiencing inbreeding depression, which in turn leads to lower fertility and a higher prevalence of genetic diseases.

Inbreeding depression is a cost that breeders impose on others, a negative externality that is a function of the number of breeders that release closely related bulls into the market. Since pedigrees transmit the externality, it is a network externality that affects all breeders and dairy farmers. Our analysis shows that the cost of this increased inbreeding ranges between 2 to 5.5 billion dollars.

A solution to the externality is yet to be found; we show that there is evidence that the increment in taits may have reached its biological maximum. Figure 6 shows that productivity PTAs may have peaked in 2015 or 2016 so that further selection does not lead to more extensive productivity growth. We also mention that the Holstein Association of the United States adjusts the values of PTAs according to inbreeding levels to penalize breeders that select bulls with high inbreeding rates to account for the higher reproduction costs of such animals. The value of this penalty is a crucial element in reducing inbreeding by incentivizing dairy farmers and animal breeders to breed lines that are less related on average by, for example, importing bulls from abroad or by crossbreeding Holsteins with other high-yield milk breeds.

References

Bourdon, R. M. (2000). *Understanding Animal Breeding*. Prentice Hall, 2 edition.

Cameron, A. C. and Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, Cambridge, UK.

Charlesworth, D. and Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, 10(11):783–796.

Ciliberto, F., Moschini, G., and Perry, E. D. (2019). Valuing product innovation: genetically engineered varieties in us corn and soybeans. *The RAND Journal of Economics*, 50(3):615–644.

Cole, J. (2019). The truth about inbreeding in dairy cattle. *Hoard's Dairyman*.

Daetwyler, H., Villanueva, B., Bijma, P., and Woolliams, J. (2007). Inbreeding in genome-wide selection. *Journal of Animal Breeding and Genetics*, 124(6):369–376.

Elsik, C. G., Tellam, R. L., Worley, K. C., et al. (2009). The genome sequence of taurine cattle: A window to ruminant biology and evolution. *Science*, 324(5926):522–528.

Falconer, D. and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Addison Wesley Longman ltd., Edinburgh Gate, Harlow, Essex CM20 2JE, England.

García-Ruiz, A., Cole, J. B., VanRaden, P. M., Wiggans, G. R., Ruiz-López, F. J., and Tassell, C. P. V. (2016). Changes in genetic selection differentials and generation intervals in us holstein dairy cattle as a result of genomic selection. *Proceedings of the National Academy of Sciences*, 113(28):E3995–E4004.

Goddard, M. and Hayes, B. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*, 124(6):323–330.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, 25(4):501–522.

Guinan, F., Wiggans, G., Norman, H., Dürr, J., Cole, J., Van Tassell, C., Misztal, I., and Lourenco, D. (2023). Changes in genetic trends in us dairy cattle since the implementation of genomic selection. *Journal of Dairy Science*, 106(2):1110–1129.

Hall, B. and Khan, B. (2003). Adoption of new technology. NBER Working Papers 9730, National Bureau of Economic Research, Inc.

Hazel, L. N. (1943). The Genetic Basis for Constructing Selection Indexes. *Genetics*, 28(6):476–490.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31(2):423–447.

Holstein Foundation (2018). Interpreting and Utilizing Haplotype Information. `https://www.holsteinusa.com/pdf/Interpreting_and_Utilizing_Haplotype_Information_1218.pdf`. Accessed: 2023-14-06.

Holstein Foundation (2021). Understanding Genetics and the Sire Summaries. `https://forms.gle/bfB57tLEijiorH7u8`. Accessed: 2023-14-06.

Hutchins, J. and Hueth, B. (2023). 100 years of data sovereignty: Cooperative data governance and innovation in us dairy. *Applied Economic Perspectives and Policy*, pages 1–26.

Katz, M. L. and Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American Economic Review*, 75(3):424–440.

Kerr, W. A. (1984). Selective breeding, heritable characteristics and genetic-based technological change in the canadian beef cattle industry. *Western Journal of Agricultural Economics*, 9(1):14–28.

Ladd, G. W. and Martin, M. B. (1976). Prices and demands for input characteristics. *American Journal of Agricultural Economics*, 58(1):21–30.

Liebowitz, S. J. and Margolis, S. E. (1994). Network externality: An uncommon tragedy. *The Journal of Economic Perspectives*, 8(2):133–150.

Lush, J. (1956). Dairy cattle genetics. *Journal of Dairy Science*, 39(6):693–694.

Lush, J. L. (1937). *Animal Breeding Plans*. Collegiate Press, Incorporated.

Makanjuola, B. O., Miglior, F., Abdalla, E. A., Maltecca, C., Schenkel, F. S., and Baes, C. F. (2020). Effect of genomic selection on rate of inbreeding and coancestry and effective population size of holstein and jersey cattle populations. *Journal of Dairy Science*, 103(6):5183 – 5199.

Melton, B. E., Colette, W. A., and Willham, R. L. (1994). Imputing input characteristic values from optimal commercial breed or variety choice decisions. *American Journal of Agricultural Economics*, 76(3):478–491.

Melton, B. E., Heady, E. O., and Willham, R. L. (1979). Estimation of economic values for selection indices. *Animal Production*, 28(3):279–286.

Smith, J. A. and Todd, P. E. (2005). Does matching overcome Lalonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1):305–353. Experimental and non-experimental evaluation of economic policy and models.

Steyn, Y., Lawlor, T. J., Lourenco, D., and Misztal, I. (2023). The importance of historically popular sires on the accuracy of genomic predictions of young animals in the US Holstein population. *Journal of Dairy Science Communiactions*, 4(4):260–264.

Wiggans, G. R. and Carrillo, J. A. (2022). Genomic selection in United States dairy cattle. *Frontiers in Genetics*, 13.

Wiggans, G. R. and VanRaden, P. M. (1989). USDA-DHIA Animal Model Evaluations. Factsheet H-2, Agricultural Research Service - USDA.

Wright, S. (1922). Coefficients of inbreeding and relationship. *The American Naturalist*, 56(645):330–338.

Yue, X.-P., Dechow, C., and Liu, W.-S. (2015). A limited number of y chromosome lineages is present in north american holsteins. *Journal of Dairy Science*, 98(4):2738–2745.