

# Quantifying Heterogeneous Returns to Adoption of Genetics:

# The Case of the Dairy Industry

Brent Hueth, Jared Hutchins, and Guilherme Rosa

Economics of Research and Innovation in Agriculture - NBER

# The Role of Biological Innovation in Dairy

- Biological innovation is a key driver of agricultural productivity (Olmstead and Rhode 2008).
- The model of the dairy sector for improving genetics was prompted by two innovations:
  - Artificial insemination (AI)  $\implies$  availability of genetics
  - Herd testing associations  $\implies$  evaluation of genetics.
- Key actors: DHIA, CDCB (formerly AIPL), AI companies.
- Result: a nearly continuous gain in dairy cow milk yield every year, 50% of which is attributed to genetics.

### How it Started



Source: AIPL

#### The Result



Source: Devries, 2017

# What About Management?

- Issue: data used for estimating mean returns for different genetics in dairy cows is **observational** data, which conflates genetic performance with mangement and selection behavior.
- Genetic technology is not selected randomly: often adopted strategically into environments or management styles where it will do well (Grilliches 1957; Suri 2011).
- What implications might this have for the productivity increases attributed to genetics in dairy? What role is "good selection" playing in increasing farm productivity?

### **Research Question**

**RQ:** Does selection behavior drive heterogeneity in returns to dairy cow genetics?

- We analyze dairy cow lactation records linked to the genetic evaluations of their sires at the time they were chosen.
- We use the Correlated Random Coefficient model as a framework for studying heterogeneity caused by input demand.
- We consider the importance of animal level selection in addition to herd level selection behavior.

# An Example Bull

TUPA	C {4}	TUPAC {4}			07/04/14   8	40 Reg. 312	<b>\$25</b>	1
Net Merit Cheese Merit Fluid Merit Daus. G Milk	CDCB PTA, AJ +\$522 69%Rel +\$528 +\$510 Herds G +1959 74%Rel	CA PTA, GENEX 8/2017 PTAT JUI''' JPI''' Fertility (SCR) PregCheck'''	+0.60 +2.9 +152 0.0 99	74%Rel 90%Rel 89%Rel	<ul> <li>Tank topping</li> <li>Unmatched C</li> </ul>	production FP		
Protein Fat CFP Prod. Life SCS LIV GI	+61 -0.04% +75 -0.08% +136 +3.0 +2.93 -0.7 41%Rel +1.3 57%Rel	HCR CCR Dtr. Pregnancy Rate EFI%	0.5 -1.5 -3.3 8.4%	50%Rel 61%Rel 62%Rel	Trait Stature Strongth Body Depth Dairy Form Rump Angle Thuri Width Rear Legs-Side Vw. Rear Legs-Side Vw. Foot Angle	Short Frail Shallow Tight Rib High Pins Narrow Posty Hock-In Low	AJCA-Irait Profile	Tall ST/ Strong 0.1 Deep NAA Open Rib 0.2 Siloped 0.2 Wide 0.1 Silckle 0.0 Silrajpht NA Staep 0.4
HARRIS X REN Sire JX SCHUL Dam FARIA BR 1-08 3056 aAa 465 DMS Beta-Casein A	IEGADE X VIBRANT TZ VOLCANO HARRIS OTHERS RENEGADE 2 d 2x 19,290m 5.1 9850 345135 1A2 Kappa-Casein A	(4) 15565 (3), VG-80% 13.6 702p lbs. B <b>BBR</b> 100			Feet & Legs Score Fore Udder Attach. Rear Udder Meight Rear Udder Width Udder Cleft Udder Oepth Front Test Piace. Rear Test Piace. Test Length Rear Test-Rear Rear Test-Rear	Low Low Low Narrow Wisk Deep Wide Short Wide Close	-1	High         High         High           Strong         0.2           High         0.4           Wide         0.2           Strong         0.3           Strong         0.3           Close         0.4           Close         0.4           Close         0.4           Close         0.4           Back         0.3           Back         0.3

#### Source: Genex

#### How to Evaluate Tupac

Tupac is evaluated with this model:

 $y = X\beta + Z\mu + e$ 

- The "animal model" explains some trait trait y as due to management X and genetics Z.
- Greatest attention is given to modeling genetic relationships to generate estimates of  $\mu$ .
- Tupac does not randomly show up in herds, however, generating bias in *μ*.

#### How to Evaluate Tupac



 Repeated choices of Tupac will update his "Predicted Transmitting Ability" μ for different traits.

#### A Production Function with Variable Returns

We model the production function of cow *i* in herd *j* at time *t* with a variable coefficient on their PTA "investment"  $z_{ij}$ :  $\mu_{ij} = \bar{\mu} + \tilde{\mu}_{ij}$ 

$$y_{ijt} = \alpha_1 + \beta_1 X_{ijt} + \mu_{ij} Z_{ij} + \epsilon_{ijt}$$
  
=  $\alpha_1 + \beta_1 X_{ijt} + (\tilde{\mu}_{ij} + \tilde{\mu}) Z_{ij} + \epsilon_{ijt}$   
=  $\alpha_1 + \beta_1 X_{ijt} + \tilde{\mu} Z_{ij} + (\tilde{\mu}_{ij} Z_{ij} + \epsilon_{ijt})$   
=  $\alpha_1 + \beta_1 X_{ijt} + \tilde{\mu} Z_{ij} + \nu_{ijt}$ 

- $E(z_{ij}\nu_{ijt}) \neq 0$ , and IV will not consistently identify  $\bar{\mu}$ .
- Correlated Random Coefficients (Wooldridge 2015) will identify under some strict linearity assumptions.

### The Econometric Model

First stage is the linear approximation of input demand using input price w, a decision made at conception date t - m:

$$z_{ij} = \alpha_{0j} + \gamma w_{t-m} + \beta_0 X_{ijt} + \eta_{ijt}$$
  
$$y_{ijt} = \alpha_{1j} + \bar{\mu} PTA_{ij} + \beta_1 X_{ijt} + \rho \hat{\eta}_{ijt} + \psi \hat{\eta}_{ijt} \times PTA_{ij} + \epsilon_{ijt}$$

- IV assumptions:
  - Assumption 1: Relevance ( $\gamma \neq 0$ )
  - Assumption 2: Independence  $(E(w_{t-m}\epsilon_{ijt}) = 0)$
- CRC Assumptions
  - Assumption 3:  $E(\nu_{ijt}|\eta_{ijt}) = \rho \eta_{ijt}$
  - Assumption 4:  $E(\tilde{\mu}_{ij}|\eta_{ijt}) = \psi \eta_{ijt}$

### The Econometric Model

$$z_{ij} = \alpha_{0j} + \gamma w_{t-m} + \beta_0 X_{ijt} + \eta_{ijt}$$
  
$$y_{ijt} = \alpha_{1j} + \bar{\mu} PTA_{ij} + \beta_1 X_{ijt} + \rho \hat{\eta}_{ijt} + \psi \hat{\eta}_{ijt} \times PTA_{ij} + \epsilon_{ijt}$$

Estimation of  $\bar{\mu}$  tells us the average effect in the data,  $\bar{\mu} + \hat{\psi}\hat{\eta}_{ijt}$  recovers the "heterogeneous effect."

- Hypothesis 1:  $\rho = 0$ , unobserved input demand unrelated to production.
- Hypothesis 2: ψ = 0, unobserved input demand does not drive heterogeneity in returns to genetics.

Including herd specific intercepts focuses specifically on  $u_{ij}$ , animal-specific match quality.

# **Collecting Data**

- DHIA data has dairy cow trait production (fat and protein) with their calving and birth dates.
- Using the CDCB website, we collected historical evaluations of dairy sires and matched them to data to know the characteristics at the time they were chosen.
- Class III component prices at time of selection are used to proxy w<sub>t-m</sub>, as they are used in the "Net Merit" index to price the traits (Vanraden et. al. 2018).

# Data Summary

Numbers of Records

	Number
Herds	3,326
Sires	10,798
Sires w/ Company ID	2,295
Dairy Cows	474,585
Number of Lactations	699,839
Lactation Records	1,660,959
State	Wisconsin
Period	June 2011 - January 2015

# Data Summary

#### Covariates

	Mean	Std Dev
Continuous Variables		
PTA Fat	28.79	27.12
PTA Protein	21.46	20.48
Proportion Milked 3x	0.59	0.49
Lactation Length	310.30	23.49
Herd Size	134.65	317.82
Binary Variables (%)		
Holstein	94.61	
Lactation Number		
1	46.03	
2	28.74	
3	15.17	
4	7.19	
5	2.87	

### Distribution of Chosen PTA's in the Data



#### **Interpreting Results**

- Theoretically, μ
   <sup>-</sup> = 1, which is perfect transmission of the trait (Kearney et. al. 2014). The effect of management with genetics is sometimes defined by how different μ
   <sup>-</sup> is from 1.
- First lactation cows and later lactation cows are analyzed separately for reasons of "survival bias."
- Robustness checks:
  - Sire company dummies
  - Alternative instruments

# **Returns to Fat**

All Lactations

	(1)	(2)	(3)	(4)	(5)
	OLS	FE	IV	IV + FE	CRC + FE
PTA Fat	0.672***	0.609***	-2.097***	-4.233***	-4.235***
	(0.0337)	(0.0138)	(0.430)	(0.422)	(0.135)
η					4.793***
					(0.135)
$\hat{\eta}$ × PTA Fat					0.00742***
					(0.0000904)
Ν	1664086	1664086	1664086	1664086	1664086
adj. R <sup>2</sup>	0.343	0.557	0.234	0.243	0.559
Standard errors in parentheses					

# **Returns to Fat**

#### Across Production Cycles

	(1)	(2)	(3)
	All Lactations	First Lactation	Later Lactations
PTA Fat	-4.235***	-2.758***	-3.136***
	(0.135)	(0.186)	(0.326)
η	4.793***	3.231***	3.779***
	(0.135)	(0.1862)	(0.326)
$\hat{\eta}$ × PTA Fat	0.00742***	0.00724***	0.00785***
	(0.0000904)	(0.000117)	(0.000139)
N	1664086	802191	861816
adj. R <sup>2</sup>	0.559	0.531	0.517

Standard errors in parentheses

# **Returns to Protein**

All Lactations

	(1)	(2)	(3)	(4)	(5)
	OLS	FE	IV	IV + FE	CRC + FE
PTA Protein	0.481***	0.406***	-0.346	-1.159***	-1.189***
	(0.0340)	(0.0113)	(0.301)	(0.227)	(0.0835)
η					1.636***
					(0.0836)
$\hat{\eta}$ × PTA Protein					0.00920***
					(0.0000934)
N	1664086	1664086	1664086	1664086	1664086
adj. R <sup>2</sup>	0.456	0.671	0.447	0.639	0.673
Standard errors in parentheses					

# **Returns to Protein**

#### Across Production Cycles

	(1)	(2)	(3)
	All Lactations	First Lactation	Later Lactations
PTA Protein	-1.189***	-1.921***	1.152***
	(0.0835)	(0.114)	(0.144)
η	1.636***	2.336***	-0.677***
	(0.0836)	(0.114)	(0.144)
$\hat{\eta}$ × PTA Protein	0.00920***	0.00895***	0.00949***
	(0.0000934)	(0.000123)	(0.000139)
Ν	1664086	802191	861816
adj. R <sup>2</sup>	0.559	0.635	0.633

Standard errors in parentheses

# **Distributions of Coefficients**



### **Robustness Checks**

- Use of alternative instruments changed the estimate for fat but not for protein; average effect is likely not pinned down for fat.
- Including sire companies as covariates did not change the results.
- Robust to using inverse hyperbolic sine transforms of all the variables.
- Finding that  $\psi > 0$  robust to all specifications.

# Conclusions

- Selection bias likely has a large effect, but effect on average returns is inconclusive.
  - Average effect here estimated as negative, but is not a robust result.
  - Modeling of traits may have to be done as a system.
- Positive selection into use of traits, at both animal and herd level.
- Biological innovation happens in the synergy between producers and the science community, not just in the lab.

Thank you!