Journal of Applied Farm Economics

Volume 4 | Issue 1 Article 5

2021

Determining the Accuracy of Holstein Bull Semen Prices with a Two-Tier Efficiency Model

Loren W. Tauer
Cornell University, lwt1@cornell.edu

Recommended Citation

Tauer, Loren W. (2021) "Determining the Accuracy of Holstein Bull Semen Prices with a Two-Tier Efficiency Model," *Journal of Applied Farm Economics*: Vol. 4 : Iss. 1, Article 5.
DOI: 10.7771/2331-9151.1060
Available at: https://docs.lib.purdue.edu/jafe/vol4/iss1/5

Determining the Accuracy of Holstein Bull Semen Prices with a Two-Tier Efficiency Model

Loren W. Tauer (Cornell University)

ABSTRACT

The pricing accuracy of dairy bull semen available for artificial insemination is investigated using a price model cast in terms of a two-tier efficiency model, which allows estimating an overprice and underprice given the net return the bull's daughter is expected to contribute to the farm based upon the transmittable trait characteristics of a bull. The estimation produces statistically significant coefficients, including the coefficients for the overpricing and underpricing of the semen. The average estimate of overpricing is around 25%, and the average estimate of underpricing is around 33%. As expected, low semen prices tend to be underpriced and high semen prices tend to be overpriced.

KEYWORDS

efficiency, genetic selection, semen price, two-tier model

INTRODUCTION

Artificial insemination of dairy cows using semen from evaluated bulls is the norm in modern milk production. This allows a farmer to select from a myriad of bulls with transmittable traits that fit into the farm production system and remediates production deficiencies of individual cows or the herd. Hundreds of bulls are available for selection with a diverse set of traits. Those traits should determine the market price of a bull's semen; bulls with higher desired traits should have higher priced semen. This paper investigates the efficiency of bull semen prices using a price model cast in terms of a two-tier efficiency model, which allows estimating an overprice and underprice given the trait characteristics of a bull embedded in a performance index.

To assist farmers in bull selection, various bull valuation measures have been developed based upon the value of traits, with values reported for each available bull. The most common of these is called the Net Merit Value developed by the USDA, which is a measure of the net profitability that a bull's daughter should contribute to the profit of the dairy farm given the traits she receives from her father.[1] If a bull produces a high Net Merit Value, and produces high profits for the dairy farm, then that bull's semen should sell for a higher price. That relationship is tested by regressing the natural log of semen prices on Net Merit Value and Net Merit Reliability allowing for random pricing error, as well as terms representing overpricing and underpricing of the semen. The results determine whether a semen price is overpriced or underpriced conditional upon Net Merit.

McGilliard (1978) published an extension type guideline for the dairy producer to determine the maximum price they can pay for profitable semen, using the proven milk production from daughters. Rogers and McDaniel (1989) investigated the impact of udder depth, teat placement, and foot angle in an index to improve breeding goals of milk yield and involuntary culling. They used selection index theory to calculate the value of inclusion of these components and found these characteristics to be of minor importance.

Richards and Jeffrey (1996) used a hedonic pricing model to identify the genetic traits of dairy bulls that determined the expected price of bull semen available in Canada. A Lifetime Profit Index in Canada, similar to Net Merit currently used in the United States, places economic values on trait characteristics based upon the direct and indirect profit these characteristics should generate in a replacement cow for a typical farm. They found that individual characteristics were better predictors of semen prices compared to the Lifetime Profit Index. The most important characteristics were milk volume, protein and fat content of the milk, with other characteristics being less important.

That is logical since the sale of milk components directly contributes to the bottom line of the profit statement, with many of the other characteristics only indirectly contributing to profit.

Schroeder, Espinosa, and Goodwin (1992) also estimated the value of genetic traits for dairy bulls using a hedonic model and found that a composite performance index and the reliability of that index performed well as significant price determinates. Although they found that the individual traits performed equally well, the inclusion of both the index and individual traits performed the best, based upon the adjusted R square values.

A risk analysis approach to rank sires was completed by Rogers (1990), who specified an expected utility function, which included not only the expected Net Merit but also an estimate of the variance of Net Merit multiplied by a risk aversion coefficient. Using different risk aversion coefficients, including risk neutrality, he showed how the ranking of a set of 375 sires changed. The rank correlation was as low as 0.82 for risk aversion measure scenarios.

Wilder and Van Vleck (1988) used data on Holstein bulls and employed linear regression to determine which traits in sires were most important in determining the price of semen. They found that regression using only the TPI (Total Performance Index) explained as much variation in semen prices as regression using specific traits, which they concluded should not be surprising given the weights on the trait components in the TPI were similar to the regression using the trait components. That supports our use of the production indexes rather than a list of traits in estimating overprice or underprice of semen given the indexes.

The purpose of this paper is not to find determinates of semen price, established by demand and supply components, but rather to determine if semen price is correctly priced given the various traits published for each dairy bull, incorporated into a net performance index. To accomplish that semen price is regressed on individual bull valuation measure using a two-tier model. The two-tier model was developed by Polachek and Yoon (1987), who applied the model to the labor market, where over- and underestimates of wage rates reflect the relative bargaining power of employers and employees. Kumbhakar and Parmeter (2009,) as well as Blanco (2017), similarly look at the labor market, and generally find employers with the bargaining power. Other applications include tourist bargaining (Zhang et al., 2017) and a real estate application by Xu et al. (2016). A recent application exploring correct pricing of U.S. Riesling wines was completed by Fried and Tauer (2019).

METHOD

The two-tier method developed by Polachek and Yoon (1987) was used to estimate an expected price function, with estimation and decomposition of the residual into three components. The first component is a normal random error. The second and third components of the residual are additional distributions above and below the expected function in addition to the normal random error, which represent overpricing and underpricing of the semen. To ensure identification separate from the normal error, these second and third components are specified as exponential distributions. Estimation is by maximum likelihood.

The equation for the i_{th} observation is:

$$y_i = x_i^{'}\delta + \varepsilon_i \tag{1}$$

where y_i is the dependent variable, x_i is the vector of covariates, δ is the corresponding parameter vector, and $\varepsilon_i = v_i - u_i + w_i$, representing the composite error term where v is a symmetric normal distribution, u an exponential distribution below the regression equation, and w an exponential distribution above the regression equation. The variable u represents underpricing and the variable w represents overpricing, controlling for the variables in the expected component of the regression equation. The assumption is that the regression shows the relationship between the expected price of the semen as a function of covariates, but that there is a normally distributed error in pricing (or data reporting), with additional errors in underpricing or overpricing.

Polachek and Yoon (1987) and Kumbhakar and Parmeter (2009) derive the probability density function of ε_{i} and the resultant log likelihood function to estimate the parameter vector δ as well as the parameters of the normal and the two exponential functions as:

$$\begin{aligned}\ln L(x;\theta) = &-n(\sigma_u + \sigma_w) \\ &+ \sum_{i=1}^{n} \ln[e^{\alpha_i}\,\Phi(\beta_i) + e^{a_i}\,\Phi(b_i)]\end{aligned} \tag{2}$$

where $\theta = \{\delta.\ \sigma_v, \sigma_u, \sigma_w)$ and

$$a_i = \frac{\sigma_v^2}{2\sigma_w^2} - \varepsilon_i/\sigma_w \quad \alpha_i = \frac{\varepsilon_i}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2} \quad b_i = \frac{\varepsilon_i}{\sigma_v} - \sigma_v/\sigma_w$$

$$\beta_i = -(\frac{\varepsilon_i}{\sigma_v} + \sigma_v/\sigma_u)$$

From these estimates the conditional distributions for u_i and w_i can be derived as:

$$E\left(\frac{u_i}{\varepsilon_i}\right) = \frac{1}{\lambda} + \exp\{\alpha_i - a_i\}\sigma_v[\varphi(-\beta_i) + \beta_i\Phi(\beta_i)]/\chi_{1i} \quad (3)$$

$$E\left(\frac{w_i}{\varepsilon_i}\right) = \frac{1}{\lambda} + \sigma_v[\varphi(-b_i) + b_i\Phi(b_i)]/\chi_{1i} \quad (4)$$

where $\lambda = \frac{1}{\sigma_u} + \frac{1}{\sigma_w}\chi_{1i} = \Phi(b_i) + \exp\{\alpha_i - a_i\}\Phi(\beta_i)$.

The following parameterizations are used if variables are included in the u and w underpricing or overpricing terms: $\mu_u = \exp(c_u + Z_i^u * B_u)$ and $\mu_w = \exp(c_w + Z_i^{w*} B_w)$, where c_u and c_w are constants, Z_i^u and Z_i^w (optional) are vectors of observational-specific variables, and B_u and B_w are the corresponding coefficient vectors. A major modeling decision is whether a covariate should be included in the expected function or in one or both of the two-tier error components. That depends upon whether the variable is thought to influence the expected function or the error around the expected function.

Because the expected values of variables u and w are conditional upon the total error, and because the expected component of the regression is estimated with error such that the regression can overlap both u and w distributions, each observation has an estimate of both overpricing and underpricing.

Finally, because the dependent variable is in logarithmic form, an exact percentage of over or underpricing is obtained as:

$$PercentOver = 100\,[e^{E(u_i/\varepsilon_i)} - 1]$$
$$PercentUnder = 100\,[e^{E(w_i/\varepsilon_i)} - 1]$$

Maximum likelihood estimates of the parameters in θ are from the software Stata using the ado file "twotier" written by Hung-Jen Wang.[2] Given that exponential functions are specified for the overpricing and underpricing errors, with a change in signs, and the random error is normally distributed, the parameters of the maximum likelihood equation (2) are identifiable (Kumbhakar & Parmeter, 2009). Those parameters are used to mathematically derive estimates of the over- and underpricing. Because the expected price equation is estimated with error v, each observation includes both an overprice estimate as well as an underprice estimate.

DATA

Holstein dairy bull genomic trait information and semen prices were obtained from the National Association of Animal Breeder certified semen services.[3] This dataset analyzed consisted of 406 active U.S. available artificial insemination (AI) Holstein bulls from the December 2018 data file for which complete data including semen prices were available. In addition to the natural log price of semen as the dependent variable, Net Merit Value and Net Merit Reliability are used as independent variables, all reported in Table 1. The price of semen is used as reported in the database with no adjustment for volume discounts or other adjustments, but is converted into natural logarithmic values because of convention in the efficiency literature to allow the underprice and overprice estimates to be percentages.

Net Merit Values are constructed from individual sire traits transmitted to daughters that reflect the economic value that daughter should contribute to farm profits over her life, using representative farm models to determine the profit impact (VanRaden, 2017).[4] For instance, two important traits with high weights in the Net Merit Value are the net value of protein and fat that a daughter will produce in a year. These are important traits because most U.S. farmers are paid for their milk based upon the pounds of protein and the pounds of fat found in the milk. Net return per pound is the price of protein or fat minus the cost of additional feed required to produce an additional pound of protein or fat, assuming other production costs remain constant with a one-pound increase in protein or fat. Other traits indirectly affect net revenue or reduce cost. For instance, a high productive life, another important trait in the Net Merit Value, reduces the cost of replacement and thus affects profit.

The published Net Merit Values are centered at zero to reflect whether the net merit of a sire is above or below peer sires. The Net Merit of the average bull in the year 2018 data was $492, with a range from negative $778 to a positive $972. Unlike the price of semen, these performance

Table 1. Summary of the Data (N = 406)

Variable	Average	Standard Deviation	Minimum	Maximum
Semen Price ($)	26	13	3	115
Net Merit (Net Return)*	492	272	–778	972
Net Merit Reliability^	91	5	76	99
Fat Pounds*	44	31	–74	121
Protein Pounds*	33	23	-69	91
Somatic Cell Count (log)*	2.8	0.2	2.4	3.4
Type (Composite)*	1.5	0.8	–1.6	4.2
Productive Life (Months)*	3.2	2.3	–6.2	8
Daughter Pregnancy Rate (%)*	0.9	2.0	–5.5	5.9
Birth Year	2012	2.3	2000	2015

* These variables are normalized around zero, where zero represents an average animal
^ Net Merit Reliability is the variance of predicted Net Merit divided by the variance of true Net Merit.
See https://aipl.arsusda.gov/reference/nmcalc-2018.htm (accessed 11/18/20).

values were not converted into natural logarithmic values given that they are constructed by normalizing around mean profit, producing many negative values. Net Merit Reliability is the variance of predicted Net Merit divided by the variance of true Net Merit. Reliability increases as more daughter information becomes available.

For robust checking, some of the important components of the Net Merit Value are used as separate variables in a separate estimation of the two-tier model. A summary of these covariates are reported in Table 1, and are again normalized around zero, to reflect potential bull performance below or above peer bulls.

All of these variables are included in the expected function rather than in either of the two-tier error terms. However, included as a variable in the two-tier error terms is the birth year of bulls to determine if younger bulls command a price premium.

RESULTS AND DISCUSSION

Estimations of the natural log of semen prices on various groups of independent variables are reported in Table 2. Model 1 has Net Merit as the sole independent variable. Nonlinearity is often used in hedonic type models. Given that Net Merit Values are centered on the mean value, producing numerous negative values, neither the log nor quadratic value of the Net Merit Value could be included in the regression. However, nonlinearity is exhibited by the natural log of the dependent variable. Model 2 includes the Net Merit Reliability in addition to the Net Merit Value, while Model 3 includes the birth year of the bull as an additional explanatory variable in the underpricing component. Model 4, for robust testing, uses as the independent variables important characteristics used in the construction of the Net Merit Value rather than the Net Merit Value itself.

All estimated coefficients in Models 1, 2, and 3 are highly statistically significant, including the coefficients for the overpricing and underpricing of the semen. The coefficient estimate of each independent variable is positive such that a higher variable value leads to higher semen prices. In the robust test equation Model 4, only the fat, somatic cell score, and cow type are statistically significant. Surprising, the protein variable is statistically insignificant. In all equations, estimates for the underpricing and overpricing coefficients are all highly statistically significant with similar estimates.

The estimates on the three error terms are all negative, but that is because the dependent variable is the natural log of the semen price. Converting to nonlog form, for instance, the random error coefficient estimate of –1.2392 in Model 1,

Table 2. Estimation of Holstein Bull Semen Prices Using Twin-Tier Models (z test statistics in parentheses)

Variable	Model 1	Model 2	Model 3	Model 4
Net Merit	0.00042*** (5.34)	0.00039*** (5.22)	0.00031*** (3.78)	
Net Merit Reliability		0.02238*** (5.33)	0.02481*** (5.77)	
Fat Pounds				0.00266*** (2.68)
Protein Pounds				0.00052 (0.38)
Somatic Cell Score				0.30653** (2.27)
Type				0.22477*** (9.03)
Productive Life				0.02503* (1.76)
Daughter Pregnancy Rate				0.01137 (0.89)
Intercept	3.0017*** (45.59)	1.0166*** (2.68)	0.84028** (2.18)	1.77787*** (4.30)
V (Random Error)	–1.2392*** (–11.93)	–1.3524*** (–11.95)	–1.3612*** (–11.94)	–1.4160*** (–11.42)
U (Underpricing)	–1.3485*** (–9.88)	–1.2704*** (–10.84)		–1.3552*** (–10.70)
U (Birth Year)			–0.7243** (–2.14)	
U (Intercept)			144.4484** (2.12)	
W (Overpricing)	–1.5639*** (–9.57)	–1.6106*** (–10.04)	–1.6298*** (–9.93)	–1.6606*** (–10.43)
Wald Chi Square	28.54***	61.87***	57.04***	125.22***
Log Likelihood	–240.07	–226.71	–224.40	–198.92
N	406	406	406	406

The symbols *, **, and *** represent statistical significance ($H_o = 0$) at the 10, 5, and 1 percent levels, respectively.

produces a value of 0.2896, the standard error of the random error.

It was hypothesized that the age of the bull may influence overpricing or underpricing of semen, such that younger bulls might be overpriced or underpriced relative to older bulls. Therefore, the birth year of bulls was first embedded into both the overpricing and underpricing variables of w and u, but the maximum likelihood function did not solve, implying that birth year was not a determinate factor in both overpricing and underpricing. The birth year was then entered only in the overpricing component and then only in the underpricing component. As shown as Model 3 in Table 2, birth year was statistically significant in underpricing such that older bulls tended to have errors in underpricing. Birth year did not affect the error in overpricing, however. It may be that demand has decreased for older bulls who may have been overused in breeding, and thus a need exists to more heavily discount their listed price. Obviously, these older bulls would be a good bargain, especially if not previously used extensively in a herd.[5]

The estimated expected relationship between the Net Merit Value and the price of semen as estimated by Model 1 in Table 2 is shown in Figure 1. The relationship, given the estimation is in semilogarithmic form, is slightly nonlinear. Also included in Figure 1 is the plot of the observations.

The derivation of the average overpricing and underpricing with standard deviation and ranges of the model estimates are reported in Table 3. The average estimate of overpricing is around 25%, and the average estimate of underpricing is around 33%. This implies that on average semen is more

Figure 1. Expected semen price conditional on Net Merit Value. Semen priced greater than $50 suppressed for graph clarity (8 observations).

Table 3. Summary Statistics of Percent Underpricing and Overpricing of Holstein Bull Semen Prices, Models 1, 2, and 4 from Table 2

Variable	Mean	Std. Dev.	Minimum	Maximum
Net Merit (Model 1)				
Overpricing	24.62	23.81	12.29	315.56
Underpricing	32.95	44.41	12.29	572.57
Net Merit and Net Merit Reliability (Model 2)				
Overpricing	21.81	33.59	7.01	397.26
Underpricing	32.95	65.04	7.01	758.27
Bull Traits (Model 4)				
Overpricing	20.74	30.27	7.31	338.03
Underpricing	29.31	54.49	7.31	686.40

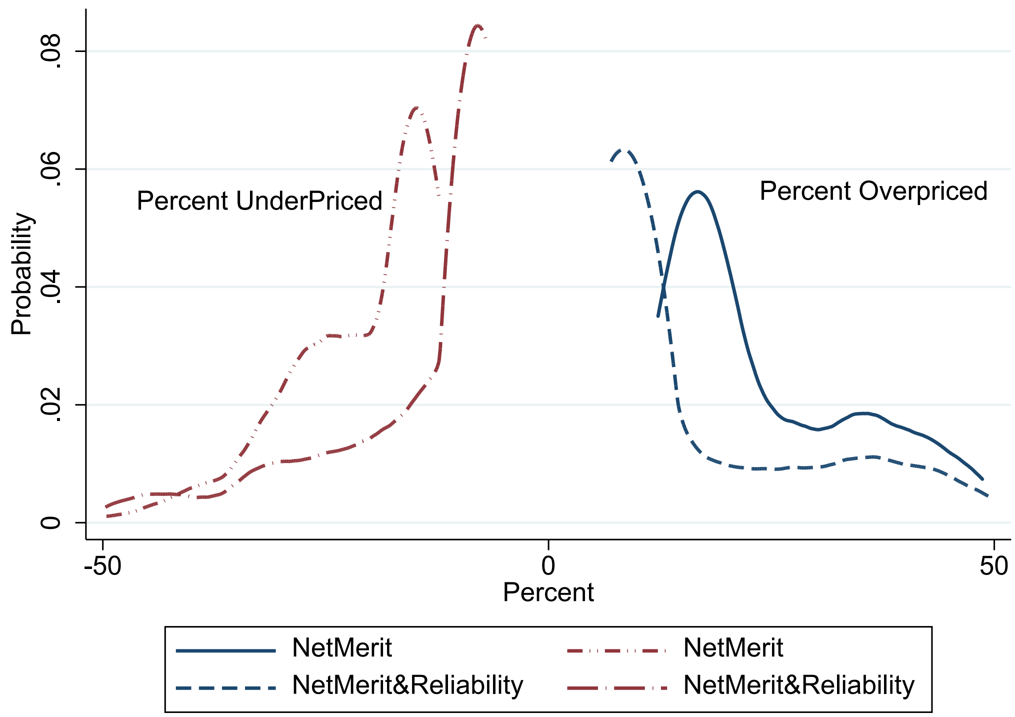

Figure 2. Percent overpriced and underpriced of Holstein bull semen using a two-tier model using the Net Merit Index alone or with the Net Merit Reliability measure.

underpriced than overpriced. Although on average overpricing and underpricing is not extreme, some semen is extremely overpriced at over 300% and other semen underpriced close to 600%. If Net Merit Value accurately reflects the value the bull will contribute to the farm, a farmer should obviously avoid semen that is overpriced and instead use the semen that is underpriced. As an alternative, when price discounts from listed prices are the norm, farmers might use the overpriced estimates to negotiate a lower price for semen. Semen sellers might also use underpriced estimates to revise listed prices.

The distribution of percentage of semen overpriced and underpriced is shown in Figure 2, which shows the results using Model 1 in Table 2. That distribution is truncated at an absolute value of 50% for best illustration of the results. However, when overprice and underprice distributions obtained from the Net Merit and Net Merit Reliability regression of Model 2 in Table 2 is also plotted in Figure 2, the amount of overpricing and underpricing both shifts to the center, reflecting the importance of Merit Reliability in pricing semen.

Low semen prices tend to be underpriced and high semen prices tend to be overpriced as illustrated in Figure 3. That relationship is inherent in pricing of any input; setting a low price increased the chance of underpricing an input, while setting a high price increased the chance of overpricing. More interesting is the pricing relationship by the Net Merit Value of the sire, as illustrated in Figure 4. There is a pattern of both underpricing and overpricing of semen as the Net Merit Value of the bull increases. However, that relationship is more pronounced with underpricing, where there is more underpricing of semen at large Net Merit Values. That pattern also appears for overpricing, but there are also overpriced bulls at low and even negative Net Merit Values.

CONCLUSION

Dairy farmers select semen to use in their herd based upon the transmittable traits of bulls and the price of the semen. Because bulls transmit many traits, a Net Merit Value measuring the overall profitability of semen to a representative herd is available to

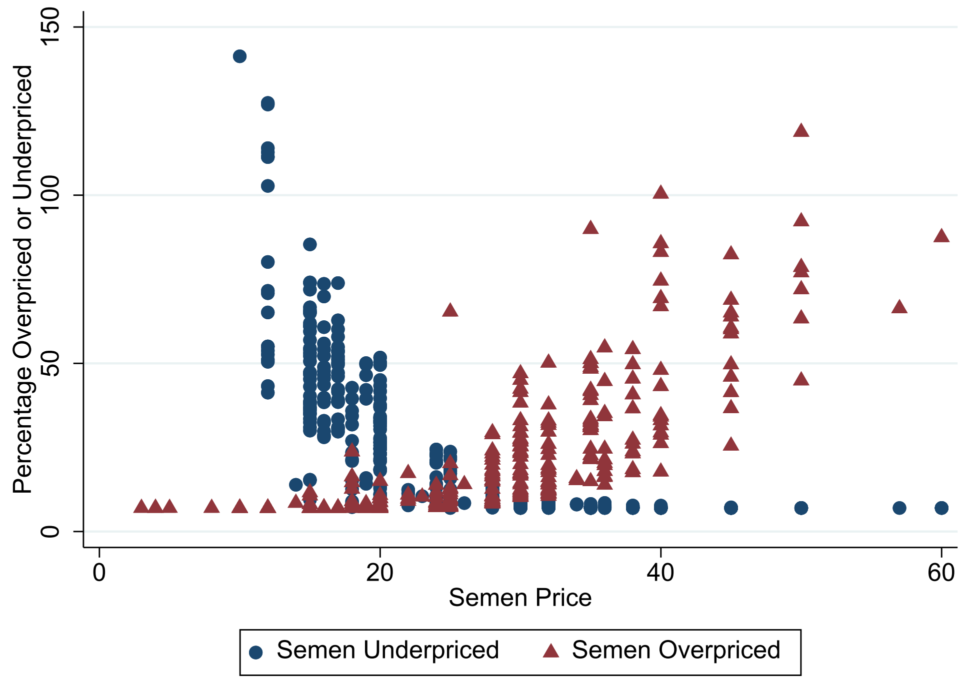

Figure 3. Percentage of overpriced or underpriced of semen prices by semen price (overprice or underprice determined by Net Merit and Net Merit Reliability two-tier regression). Four underpriced outliers, three overpriced outliers, and four semen prices greater than $75 were removed for graph clarity.

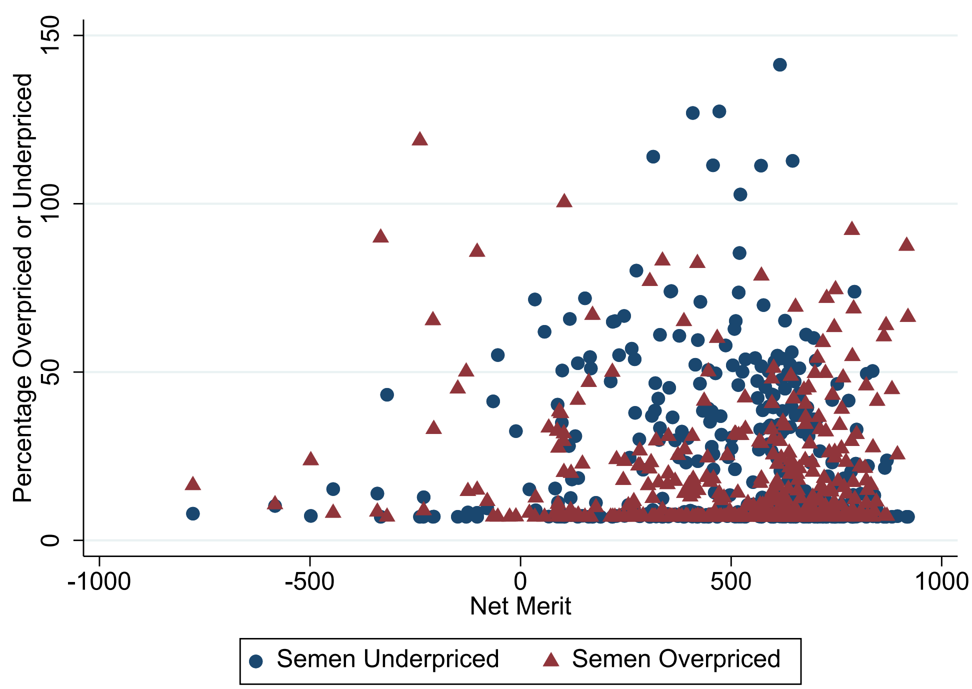

Figure 4. Percentage of overpriced or underpriced of semen prices by Net Merit (overprice or underprice determined by Net Merit and Net Merit Reliability two-tier regression). Four underpriced outliers, three overpriced outliers, and four semen prices greater than $75 removed for graph clarity.

help in this selection. This statistic represents the profitability that the various bull-transmitted traits would instill into their daughters. As such, there should be a relationship between the price of semen and the value of that semen on the farm.

The two-tier model was used to estimate the relationship between the price of semen and Net Merit Value, with estimates of overpricing and underpricing of semen. An estimate of average overprice is 25%, and the average underprice is greater at 33%. Farmers can use the results to determine if a specific bull semen is a good buy and use the results to negotiate a lower price. Semen producers can also use the results for repricing semen. Because new bulls are continuously added and other bulls are deleted, the results can be updated periodically.

ACKNOWLEDGMENTS

Loren Tauer is a faculty member in the Charles H. Dyson School of Applied Economics and Management, Cornell SC Johnson College of Business, and College of Agriculture and Life Sciences, Cornell University. This research was funded using Cornell faculty funding.

NOTES

1. Other Merit estimates include Fluid, Cheese, and Grazing, for farms that produce fluid milk priced by weight, farms producing cheese with protein more highly valued, and grazing farms where preferred cow type is different. Correlation coefficients of these alternative measures with the Net Merit Value were all 0.99 or higher, precluding the necessity of estimating with these alternative values.

2. The book by Kumbhakar and Wang (2015) discuss the two-tier frontier method and Jung-Jen Wang provided the code.

3. https://www.naab-css.org/databases

4. The (2018) USDA Net Merit is constructed using the following traits and weights: Protein (17%), pounds of Fat (27%), Milk Yield (–1%), Productive Life (12%), Somatic Cell Score (-4%), Udder Composite (7%), Feet/legs Composite (3%), Body Size Composite (-5%), Daughter Pregnancy Rate (7%), Heifer Conception Rate (1%), Cow Conception Rate (2%) and Calving Ability Dollars (5%).

5. Given the technology of producing semen for insemination, a single bull is used to service many thousands of cows. Given that semen is frozen for storage, a bull's semen can be used years after his death.

REFERENCES

Blanco, G. (2017). Who benefits from job placement services? A two-sided analysis. *J. Productivity Analysis, 47*, 33–47.

Fried, H. O., & Tauer, L. W. (2019). Efficient wine pricing using stochastic frontier models. *J. of Wine Econ., 14*, 164–181.

Kumbhakar, S. C., & Parmeter, C. F. (2009). The effects of match uncertainty and bargaining on labor market outcomes: Evidence from firm and worker specific estimates. *J. Productivity Analysis, 31*, 1–14.

Kumbhakar, S. C., & Wang, H.-J. (2015). *A practitioner's guide to stochastic frontier analysis using Stata*. Cambridge University Press.

McGilliard, M. L. (1978). Net returns from using genetically superior sires. *J. Dairy Sci., 61*, 1680–1682.

Polachek, S. W., & Yoon, B. J. 1987. A two-tier earnings frontier estimation of employer and employee information in the labor market. *Review Econ. and Stat.*, *69*, 296–302.

Richards, T. J., & Jeffrey, S. R. (1996). Establishing indices of genetic merit using hedonic pricing: An application to dairy bulls in Alberta. *Canadian J. Agri. Econ., 44*, 251–264.

Rogers, G. W. (1990). A utility function for ranking sires that considers production, linear type traits, semen cost, and risk. *J. Dairy Sci.*, *73*, 532–538.

Rogers, G. W., & McDaniel, B. T. 1989. The usefulness of selection for yield and functional type traits. *J. Dairy Sci., 72*, 187–193.

Schroeder, T. C., Espinosa, J. A., & Goodwin, B. K. (1992). The value of genetic traits in purebred dairy bull services. *Review Agri. Econ*., *14*, 215–226.

VanRaden, P. M. (2017). Net merit as a measure of lifetime profit: 2017 revision. *USDA*. https://www.aipl.arsusda.gov/reference/nmcalc-2017.htm#Derivation of economic values

Wilder, J. S., & Van Vleck, L. D. (1988). Relative economic values assigned to milk, fat test, and type in pricing of bull semen. *J. Dairy Sci*., *71*, 492–497.

Xu, H., Wang, H., Zhou, C., & Geng, Z. (2016). Dual attributes of government intervention and China's real estate prices—Is Chinese government a promoter of the real estate prices. *Accounting and Finance Research*, *5*, 29–36.

Zhang, H., Zhang, J., Yang, Y., and Zhou, Q. (2017). Bargaining power in tourist shopping. *J. Travel Research, 57*, 947–961.