A review of population data utilization in beef cattle research\textsuperscript{1}

R. Jones*\textsuperscript{2} and M. Langemeier†

*NW Area Extension Office, Oklahoma State University, Enid 73701; and †Department of Agricultural Economics, Kansas State University, Manhattan 66506

ABSTRACT: Controlled experimentation has been the most common source of research data in most biological sciences. However, many research questions lend themselves to the use of population data, or combinations of population data and data resulting from controlled experimentation. Studies of important economic outcomes, such as efficiency, profits, and costs, lend themselves particularly well to this type of analysis. Analytical methods that have been most commonly applied to population data in studies related to livestock production and management include statistical regression and mathematical programming. In social sciences, such as applied economics, it has become common to utilize more than one method in the same study to provide answers to the various questions at hand. Of course, care must be taken to ensure that the methods of analysis are appropriately applied; however, a wide variety of beef industry research questions are being addressed using population data. Issues related to data sources, aggregation levels, and consistency of collection often surface when using population data. These issues are addressed by careful consideration of the questions being addressed and the costs of data collection. Previous research across a variety of cattle production and marketing issues provides a broad foundation upon which to build future research. There is tremendous opportunity for increased use of population data and increased collaboration across disciplines to address issues of importance to the cattle industry.

Key words: cost, economic analysis, efficiency, financial data, production data, profit

INTRODUCTION

Without a doubt, controlled experimentation and analysis of the resulting data using sample data analysis techniques will continue to provide the preferred protocol for most of the basic research done in the physical sciences. However, alternative methods that make use of data obtained from sources other than controlled experimentation, such as population data, warrant consideration and may increase in usage in the future for a variety of reasons. First, the ever increasing cost associated with controlled experimentation combined with tighter research budgets means the entire research community will strive to increase the efficiency of the research process. Second, there are already several population data sets available, both public and private, that could be used to a greater degree than they are being used currently to help address livestock production and marketing issues. Finally, many important research questions, particularly those that are multidisciplinary in nature, lend themselves to analysis using alternative methods. This is certainly the case when combining physical production questions with social science questions, such as combining an economic analysis with an animal science production question. In this example, applied economists should be well positioned to contribute to such joint research because analysis using population statistics is the norm rather than the exception in much applied economic research.

The purpose of this paper is to provide additional insight regarding the use of population statistical evaluation techniques for use in analyzing animal production systems. It is important that researchers, especially young professionals such as graduate students across the spectrum from biological scientists to social scientists, understand the contributions that various approaches make to the furtherance of science and gain a better understanding of specific circumstances in which particular approaches may prove useful. In this paper we first provide a discussion of some alternative analysis methods that have been used to examine animal

\textsuperscript{1}Based on a presentation at the Beef Symposium titled “Population data analyses to evaluate trends in animal production systems” at the Joint Annual Meeting, July 12 to 16, 2009, Montreal, Canada. Manuscript publication sponsored by the Journal of Animal Science and the American Society of Animal Science.

\textsuperscript{2}Corresponding author: rodney.jones@okstate.edu

Received July 30, 2009.
Accepted October 25, 2009.
production and marketing issues. We then briefly address data issues, such as type and quality. Finally, to provide examples, we briefly review several different types of studies that have used or are using population data analysis methods to examine issues relevant to livestock production and marketing.

DISCUSSION

Statistical Methods

Although a wide array of statistical methods can be utilized for analysis of population data in studies of interest to animal production and management, the vast majority of work involves either some variation of statistical regression or some form of mathematical programming. Therefore, this discussion focuses on these 2 methods. The most common tool used for population data analysis for the last several years has been some form of statistical regression, where some variable of interest is modeled as a function of a set of other variables. Examples relevant to animal agriculture include BW gain being modeled as a function of physical or environmental factors (e.g., Mark et al., 2000), or profit being modeled as a function of various physical factors along with prices or other external economic factors (e.g., Schroeder et al., 1993b).

In the most basic form, the classical linear regression model rests on some basic assumptions about the way in which observations are generated (Kennedy, 1992). The first assumption is that the dependent variable can be calculated as a linear function of a specific set of independent variables. The second assumption is that the expected value of the error term across all observations is zero. The third assumption is that the distribution of the error term does not change across observations of the various independent variables. The fourth underlying assumption is that the observations on the independent variable can be considered fixed in repeated sampling. The final assumption, which is a mathematical requirement for estimation, is that the number of observations is greater than the number of independent variables and that there are no exact linear relationships between the independent variables.

As the researcher deviates from the world of tightly controlled experimentation, it becomes more likely that there will be severe violations of the underlying assumptions of the classical linear regression model, which if not accounted for can lead to incorrect scientific conclusions (e.g., Kennedy, 1992). Fortunately, statisticians, social scientists, and others who rarely have the luxury of controlled experimentation or exceptionally clean data have come up with a long list of sophisticated statistical techniques for which to test and deal with problems caused by violations of the standard assumptions. In the real world of applied analysis, it is almost always a matter of degree. The scientist accepts the fact that one or more of the underlying assumptions will be violated and employs a battery of tests to determine the degree to which each assumption is violated. Subsequently, a determination is made regarding the need to employ a correcting technique.

In addition to using regression analysis, it is common to use nonparametric efficiency analysis to examine technical and cost efficiency. Nonparametric efficiency analysis utilizes mathematical programming to find the production and cost frontier (Fare et al., 1985). Efficiency studies are motivated by the large differences in performance among farms, which are documented by numerous economic studies (e.g., Yeager and Lange-meier, 2009).

Efficiency analysis may involve an examination of technical efficiency, cost efficiency, or both. Farms, ranches, or other firms that are technically efficient produce on the production frontier. These business entities are producing the greatest level of output for a given level of inputs. Firms that are technically inefficient could expand output with the same level of inputs by improving their technical efficiency. Farms, ranches, or other firms that are cost efficient produce on the cost frontier. Cost efficiency is also referred to as economic or overall efficiency. These business entities are producing at the least cost for a given level of output. Firms that are cost inefficient could reduce costs for a given level of output by improving cost efficiency.

It is common for studies to combine efficiency analysis techniques with regression analysis (e.g., Featherstone et al., 1997). Efficiency measures, such as technical efficiency or cost efficiency, are computed for each observation and then regressed on other variables of interest. These factors could include demographic variables, such as age, time trend, or weather variables. The objective is to learn about various factors that contribute to efficiency at the farm level.

Data Requirements

Data requirements for analysis using controlled experimentation are fairly well understood. However, when operating in the world of population data analysis, several data-related questions arise for debate and discussion. One of the most common debates centers around the level of data aggregation that is appropriate for the analysis at hand. In other words, do we need observations associated with each individual animal, or will averages across animals in a pen or associated with a particular farm or firm be adequate? An additional aggregation debate centers around the time period dimension, such as data representing days, weeks, months, and so on. These become relevant issues because certain data may already be available from public or private sources, and the cost of collecting data at a different aggregation level can become quite high. Therefore, the question often becomes one of “can I answer the question at hand with data that are already available or with data that can be acquired at a low enough cost?” The acceptable level of aggregation for a particular study depends on the questions be-
data are costly, resulting in the ongoing debate about a uniform straight line formula or it may be calculated. Depreciation expense may sometimes be calculated based on a consistent manner. For example, depreciation performed in a consistent manner. For example, depreciation may be calculated based on the data have been measured and collected in a uniform manner. For example, ADG at the pen level could be measured with additional calculations performed on the data that might contribute to feed conversion for a particular observation, such as differences across farms, weather, and so on. An analysis looking at the financial efficiency of a particular production practice on a particular production unit would require detailed input level and cost data, along with a physical output measure or a value of production measure. Quite often adjustments are needed to account for inventory changes or output that is not the primary output of interest, such as cull cow sales in an analysis of calf output.

Across all levels of aggregation and specific data types, consistency in collection and calculation becomes more of an issue when using population data for analysis when compared with the controlled experiment protocol. The reason is obvious. In the controlled experiment the researcher or research team provides oversight in the data collection process and dictates any subsequent calculations uniformly across observational units. When using population data collected across observational units, such as farms or other business entities over time, one must ensure that the data have been measured and collected in a uniform manner. For example, ADG at the pen level could be measured with dead animals included on 1 farm and not included on another. Also, the researcher must ensure that any additional calculations performed on the data have been performed in a consistent manner. For example, depreciation expense may sometimes be calculated based on a uniform straight line formula or it may be calculated and reported based on inconsistently applied tax management rules. Obviously, assuring consistency in the data are costly, resulting in the ongoing debate about quality vs. quantity. The bottom line is that the researcher must ensure that the data are of sufficiently high quality to answer the questions of importance.

There may well be instances when the cost of acquiring the ultimate level of detail in a particular collection effort becomes prohibitively high, and the researcher is resolved to addressing a different level of the question at hand due to the realities of the data collection process.

A lot of research is conducted using a combination of population data pertaining to some physical production process along with a set of publically acquired general data on prices, weather, or other measures. Doing so allows the researcher to examine the influence of outside factors on farm or industry performance. Several important issues of interest to the livestock industry have been addressed in this manner. As an example, physical production data are often combined with publicly acquired output price and input cost data to construct enterprise budgets for various comparison purposes. The appropriate time dimension of the price data may depend on the purpose of the economic comparison. If the intent is to compare alternative production systems over a long time period of time, then it may be perfectly appropriate to combine physical performance data from a short-term study with long-term average price and cost data. Five- to ten-year averages are commonly used to represent long-term agricultural prices. In other cases, the researcher may want to combine production data from the most recent physical experiment that is relevant, which may be several years old, with contemporary price data or even future price projections to obtain an enterprise budget projection for a particular year or production cycle. Alternatively, the researcher may be interested in adding an economic component to a specific physical experiment for a specific time period, in which case one must match up the time dimension of the observational units between the physical production process and the data acquired from other sources.

There are several examples of excellent data sources that have been or could be used to address issues related to the livestock industry. Beginning with the broadest and most readily available, there are obviously publicly available data sources of prices of livestock, prices of meat products, feed costs, and broad measures pertaining to input costs for livestock production. These data have often been combined with data from specific pens of cattle, for example, or from specific farms to study farm performance or the impact of outside influences on farm performance. Various sources of feedlot performance data representing various levels of data aggregation are available and have been used to study trends, seasonality, and other cattle feeding issues of interest. Similar data are being collected for other species, but are perhaps less readily available. Detailed farm-level data that represent production, specific costs, and specific revenue flows are available from a few sources and are being used on an ongoing basis to evaluate factors that impact individual farm performance. Examples include data that are collected and maintained by Farm Management Associations, such as the Kansas Farm Management Association network, and the data that
have been collected over the years though the Integrated Resource Management–Standardized Performance Analysis efforts. In the opinion of the authors, these data are perhaps the most useful and informative, but are also the most costly to collect and maintain.

**Examples of Previous Studies**

A wealth of previous and ongoing research has utilized population data to address issues of importance to the beef industry. Here, we provide a review of a few examples with which we are familiar, primarily with the intent of exposing the reader to the variety of types of issues that can potentially be addressed using these techniques. The works referenced herein represent only a small sample of the body of research that has been done and is ongoing.

One significant body of relevant research addresses marketing and pricing issues in the beef industry. Industry participants are well aware of the ongoing debates regarding market structure at various levels of the beef marketing chain combined with the various marketing tools that are utilized, and the resulting market performance in terms of price levels, potential for market manipulation, and so on. The results of previous studies addressing such issues have varied. For example, early research using pen-level transaction data combined with aggregate measures of industry market behavior indicated that the level of forward contracting activity, or captive supply, had only minor impacts on transaction prices in the residual cash markets for fed cattle, and could have positive impacts on the overall prices received for all cattle transacted over a given time period (Schroeder et al., 1993a; Morrison-Paul, 2001). Using similar population data, other researchers have concluded that cash market price distortions could be significant (Crespi and Sexton, 2005). The results of such market performance studies can be extremely sensitive to the specific models employed, as well as to the level of data aggregation (Jones et al., 1996). Recent research offers further insight regarding why previous studies have reached such differing conclusions, linking the degree of market price distortions to the phase of the cattle cycle during the period of observation or data collection (Crespi et al., 2010).

Several studies over a period of several years have examined the impact of individual animal characteristics on transaction prices. For example, Jones et al. (1992) used pen-level cash market transaction data to examine the impact of various quality characteristics, such as breed, hide color, estimated quality grade, or estimated yield grade, on fed cattle transaction prices. Others have used sale barn or other transaction level observations to evaluate the impact of feeder cattle characteristics on transaction prices (e.g., Schroeder et al., 1988; Lambert et al., 1989; Turner et al., 1991). A more recent extension of this body of research examines the value of management alternatives such as preconditioning and identification or source verification for feeder cattle and calves (Avent et al., 2004; Bailey, 2004; Dhuyvetter et al., 2005; Dickinson and Bailey, 2005). Still others, such as Mintert et al. (1990), Parcell et al. (1995), and Jones et al. (2008), have used combinations of transaction level data and aggregate market information to look at the contribution of individual characteristics to breeding animal value. Studies of this type are not only interesting from a research perspective, but extremely useful to industry participants in helping to understand the magnitude of these price influencing factors and how they may have changed over time and across geographic locations.

An ongoing body of research continues to examine the factors that contribute to production and economic risk in the cattle feeding industry. Over the years, several studies documented the seasonality of feedlot performance and the relative contributions of various factors to performance and ultimately to feeding cost of BW gain (e.g., Langemeier et al., 1992; Schroeder et al., 1993b; Mark et al., 2000). Studies of this type help feeders make informed cattle procurement and placement decisions and assist in the development of sound economic risk management strategies. Over time, as conditions, technology, and feeding practices evolve, it is important that studies like these be updated and refined so that producers can make the most informed cattle feeding decisions. Somewhat related, but from a longer term industry perspective, several researchers (e.g., Langemeier and Jones, 2000; Langemeier et al., 2001) have examined the pace of technological improvement in cattle feeding compared with other industries, using detailed performance and input cost data. Results of such studies, although sometimes discouraging, provide direction and motivation for continued production research to improve livestock production technology so that the livestock sector remains competitive.

Several early applied economic researchers (e.g., Bradford et al., 1978; Buccola et al., 1980; Musser et al., 1980) examined various economically important issues pertaining to the backgrounding phase of the cattle industry. These studies reinforced the importance of examining individual farm-level decision units when addressing issues pertaining to the economics of this phase of the industry. More recently, several studies have utilized population data to examine retained ownership management issues or issues of interest to the economics of this phase of the industry. For example, Wang et al. (2001) looked at how production and price risk factors combine with individual risk preferences to assist cow-calf producers in the decision to retain ownership. White et al. (2007) looked at how historical farm-level data or university feed-out data can be combined with grid pricing data to assist in retained ownership and marketing decisions.

Numerous studies have used statistical analysis techniques applied to population data samples to examine the interactions between various animal health issues and economically important outcomes. For example, Ir-sik et al. (2006) quantified the impact of common ani-
nal health issues on the performance of feedlot cattle, and Perry et al. (2007) examined management strategies that jointly impact the occurrence of Escherichia coli O157:H7 and feedlot profits. Pendell et al. (2007), as well as others, have explored the impact of a broad-based disease outbreak such as foot and mouth disease on the livestock industry as well as on the broader economy. Again, studies such as these help industry participants make intelligent choices regarding cattle procurement and disease prevention and management protocols.

The issue of international trade for both live cattle and for beef has been a controversial topic for many years in the cattle industry. Although individual farm-level data are rarely needed to examine trade-related issues, certainly data acquired from population sampling have been used to evaluate various trade issues. Marsh et al. (2005) provide an example of an evaluation of the impacts, both positive and negative, of easing trade restrictions on beef producers. As an example of a more specific trade issue that has received considerable research attention, several researchers including, but not limited to, Brester et al. (2004) and Lusk and Anderson (2004) have examined the impacts of country-of-origin labeling and related policies on industry participants and consumers.

Examples of efficiency studies include Featherstone et al. (1997) and Langemeier and Jones (2006). Featherstone et al. (1997) examined technical and cost efficiency for a sample of Kansas beef cow enterprises. On average, the farms were 78% technically efficient and 60% overall efficient. Regression analysis was used to relate efficiency estimates to input expenses and farm characteristics. Feed, labor, and capital expenses contributed to cost inefficiency. Cost efficiency was positively related to herd size and enterprise profitability. Langemeier and Jones (2006) examined scope and cost efficiency for a sample of Kansas beef cow enterprises. On average, the farms were 78% technically efficient and 60% overall efficient. Regression analysis was used to relate efficiency estimates to input expenses and farm characteristics. Feed, labor, and capital expenses contributed to cost inefficiency. Cost efficiency was positively related to herd size and enterprise profitability.

SUMMARY AND CONCLUSIONS

A wide array of questions of interest to beef cattle production scientists can be addressed using population data and appropriate analysis techniques. This presents a tremendous opportunity for increased collaboration between biological scientists and social scientists such as applied economists. However, usable population data are not costless. Studies typically require data from more than one firm or collection unit (e.g., farm, feedlot location, and so on), involving a considerable amount of time and effort. In addition, there must be assurance that measures are consistently collected so that they represent the same thing across observations. With that said, there are a lot of good data sets that are already being collected for other purposes that could be more fully utilized for research to address issues of interest to the animal production industry.

LITERATURE CITED


