

MODELLING ANIMAL SYSTEMS PAPER

Modelling semi-continuous data using mixture regression models with an application to cattle production yields

E. J. BELASCO¹* AND S. K. GHOSH²

¹ Department of Agricultural and Applied Economics, Texas Tech University, Lubbock, TX, USA

² Department of Statistics, North Carolina State University, NC, USA

(Received 2 September 2010; revised 17 May 2011; accepted 21 June 2011; first published online 20 July 2011)

SUMMARY

The present paper develops a mixture regression model that allows for distributional flexibility in modelling the likelihood of a semi-continuous outcome that takes on zero value with positive probability while continuous on the positive half of the real line. A multivariate extension is also developed that builds on past multivariate models by systematically capturing the relationship between continuous and semi-continuous variables, while allowing for the semi-continuous variable to be characterized by a mixture model. The flexibility associated with this model provides potential applications in many production system studies. The empirical model is shown to provide a more accurate measure of mortality rates in cattle feedlots, both independently and within a system including other performance and health factors.

INTRODUCTION

Semi-continuous dependent variables have long been a challenging modelling problem associated with disaggregated data sets. The most common occurrences are found in consumption and production data. Regarding consumption, households typically do not purchase all the goods being evaluated in every time period, resulting in observed zero values with positive probability. Similarly, a study evaluating the number of defects in a given production process is likely to have outcomes with no defects, leading to a high proportion of observed zeros. In both cases, ordinary least squares parameter estimates will be biased when applied to these types of data (Amemiya 1984).

Tobin (1958) was the first to recognize this bias and offered a solution that is still quite popular today. However, the assumption of normality of the underlying latent variable has made the Tobit model inflexible to data generating processes outside of that major distribution (Bera *et al.* 1984). Additionally, Arabmazar & Schmidt (1982) demonstrated that random variables modelled by the Tobit model contain substantial bias when the true underlying distribution

is non-normal and has a high degree of censoring. The Tobit model has been generalized to allow variables to influence the probability of a non-zero value and the non-zero value itself as two separate processes (Cragg 1971; Jones 1989), which are commonly referred to as the Hurdle and double-Hurdle models, respectively (another model that allows for decisions or production output processes to be characterized as a two-step process is the zero-inflated class of models).

The univariate Tobit model is extended, under a set of mild assumptions, to include multivariate settings (Amemiya 1974; Lee 1993). While empirical applications in univariate settings are discussed by Amemiya (1984), multivariate applications are becoming more frequent (Cornick *et al.* 1994; Chavas & Kim 2004; Eiswerth & Shonkwiler 2006; Belasco *et al.* 2009a). Shonkwiler & Yen (1999) provide a more general procedure to consistently estimate a system of equations under fewer restrictions than the multivariate Tobit model. However, in their proposed model, parameter estimates and standard errors are not systematically estimated, nor are cross-correlation between equations or heteroskedasticity explicitly modelled. An alternative that allows for more distributional flexibility was first derived by Duan *et al.* (1983) who, in an evaluation of medical expenses, assumed a

* To whom all correspondence should be addressed. Email: eric.belasco@montana.edu

two-step process where a decision is first made regarding positive expenses, followed by a second decision about how much to spend, conditional on a positive amount of expenses.

The present study considers the use of a mixture model that extends Duan *et al.* (1983) to characterize semi-continuous dependent variables, in univariate and multivariate situations. This model will be shown to nest the Tobit model, while major advantages include the flexibility in distributional assumptions and an increased efficiency in situations involving a high proportion of zero values. Additionally, the multivariate version allows for an explicit characterization of the dependence between the semi-continuous and continuous dependent variables. For the empirical study, a semi-continuous log-normal mixture (SLNM) regression model is derived from a generalized mixture regression model. A comprehensive dataset is used that includes proprietary cost and production data from five cattle feedlots in Kansas and Nebraska, amounting to 11 397 pens of cattle during a 10-year period. Cattle mortality rates (MORT) on a feedlot provide valuable insights into the profitability and performance of cattle on feed. Additionally, it is hypothesized that cattle MORT are more accurately characterized with an SLNM regression model that takes into account the positive skewness of MORT, as well as allowing observations from discrete and continuous portions of the semi-continuous distribution to be modelled separately. In both univariate and multivariate situations, the proposed SLNM regression model more efficiently characterizes the observed features of the data. The objective of the present research is to develop an alternative strategy to modelling semi-continuous data that can more effectively characterize cattle MORT alone as part of a system of equations that includes indicators of cattle productivity.

Section 2 develops a generalized mixture regression model that is specified with a log-normal distribution and is used for estimation in the present research. The univariate model precedes the development of a multivariate model. Repeated simulations are used to evaluate the penalty, in terms of in-sample and out-of-sample model fit, of assuming the wrong model. Next, the proposed model is applied to evaluate data from commercial cattle feedlots in Kansas and Nebraska. Comparisons between the Tobit and mixture regression model are assessed through model fit, prediction power and parameter estimates using both univariate and multivariate models using simulated

and empirical data. The final section provides the implications of the study and avenues for future research.

MATERIALS AND METHODS

A semi-continuous variable Y is assumed that can take on the value of zero with a positive probability or a positive value and is conditional on variables contained in X . Shonkwiler & Yen (1999) assume the following set of equations:

$$Y_i = W_i \times (G(X_i, \beta) + \varepsilon_i) \quad (1)$$

$$W_i = I(H(X_i, \gamma) + v_i > 0) \quad (2)$$

where the residual errors ε_i 's and v_i 's are distributed according to a bivariate normal. The binary-valued variables W_i 's measure the probability of zero or positive outcomes, which are commonly characterized in the literature using a Bernoulli variable. Then, the positive outcomes are modelled separately as realizations of positive valued random variables V_i . The regression parameters β and γ are vectors that include marginal impacts to both W_i and V_i , respectively. Duan *et al.* (1983) make a slight modification to Eqn (1) to make it a conditional regression with a logged dependent variable in order to account for positive skewness. This flexibility comes from the fact that the likelihood function is multiplicatively separable and ε and v do not rely on independence or distributional assumptions. Since Shonkwiler & Yen (1999) use an unconditional response in the second step, the residuals in Eqns (1) and (2) are assumed to be from a bivariate normal distribution.

A major difference between the mixture and Tobit model is that the parameters of the distribution of the unobservable censored observations are modelled separately. The probability density function of the semi-continuous random variable Y can be expressed as follows:

$$f(y|\theta) = \begin{cases} 1 - \rho(\theta), & \text{if } y = 0 \\ \rho(\theta)g(y|\theta), & \text{if } y > 0 \end{cases} \quad (3)$$

where $g(y|\theta)$ is a probability density function satisfying $\int_0^\infty g(y|\theta) dy = 1$ and $\rho(\theta) \in [0, 1] \forall \theta$. This is equivalent to the fact that $W \sim \text{Ber}(\rho(\theta))$ and $V \sim g(y|\theta)$, where W and V are assumed to be independent and $Y = WV$. This formulation includes the standard univariate Tobit model when $\theta = (\mu, \sigma)$, $\rho(\theta) = \Phi(\mu/\sigma)$, and

$$g(y|\theta) = \frac{(1/\sigma)\phi((y - \mu)/\sigma)}{\Phi(\mu/\sigma)} I(y > 0)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$, respectively, denote the density and distribution functions of a standard normal distribution. Note that in the log-normal specification to follow, ρ is modelled separately, using a set of parameters that may or may not be used to parameterize the mean and variance of V , making it more flexible than the Tobit model. The above formulation includes the models proposed by Shonkwiler & Yen (1999) and that of Duan *et al.* (1983), if dependence between W and V is allowed. Their models have been extended in the present paper by allowing the residual errors to depend on the predictors X_i 's to capture the possible heteroskedasticity.

Next, a univariate mixture regression model is developed that includes predictor variables, which then can be extended to allow for multivariate cases. Since only the positive outcomes were modelled through the second component, the natural log of the dependent variable was used. In most cases, taking the natural log of this variable works to make the dependent variable that was originally positively skewed approximately symmetrical, in addition to approximately stabilizing the variance. Using a log-normal distribution for the random variable V and allowing ρ to vary based on the conditioning variables, the basic mixture regression model can be transformed into the following form that can be generalized to include continuous distributions. First, the normal distribution is derived to model the logarithm of the positive-valued dependent variable outcomes, also known as the log-normal distribution, of the following form:

$$f(y_i|\beta, \alpha, \gamma) = 1 - \rho_i(\gamma), \quad \text{if } y_i = 0$$

$$= \rho_i(\gamma) \frac{1}{y_i \sigma_i(\alpha)} \phi\left(\frac{\log(y_i) - x_i\beta}{\sigma_i(\alpha)}\right), \quad \text{if } y_i > 0$$

(4)

where $\phi(\cdot)$ denotes the probability density function of a standard normal distribution with mean zero and variance unity; β, α and γ are $(K \times 1)$ vectors of regression coefficients; x_i is a $(1 \times K)$ vector of predictor variables; $\rho_i(\gamma) = 1/(1 + \exp(x_i\gamma))$ is the logit link function and $\sigma_i^2(\alpha) = \exp(x_i\alpha)$ is the conditional variance that guarantees $\sigma_i^2 > 0$ and $\rho_i \in (0, 1) \forall i$. K is equal to the number of conditioning predictor variables or covariates. Note that this specification is nested within the generalized version in Eqn (3), where $g(y|\theta)$ is a log-normal distribution and $\theta = (\gamma, \beta, \alpha)$. This model will be denoted by SLNM (ρ, μ, σ^2) , where ρ denotes the link function (e.g. logit, probit, etc.), μ denotes the mean function on the logarithmic scale and σ^2 denotes

the variance function on the logarithmic scale of the positive part of the dependent variable. Other candidate distributions for V 's were also assessed, including the Gamma and normal distributions, but were found to be inferior in terms of model fit and prediction accuracy for the present datasets.

To model multiple dependent variables in a way that captures the cross-correlation between the response variables, the well-known relationship between the joint and conditional marginal densities was used. More specifically, the well-known decomposition of a joint density function was used as the product of conditional and marginal densities; $f(y_1, y_2) = f(y_1|y_2) \times f(y_2)$ (Casella & Berger 2001) to capture the dependence between Y_1 and Y_2 , where Y_1 is a semi-continuous variable that has a positive probability of taking on the value of 0 and Y_2 contains the remaining vectors of continuous dependent variables. In this case, $f(y_1, y_2)$ denotes the joint density function of (Y_1, Y_2) , $f(y_1|y_2)$ is the conditional density of Y_1 , given Y_2 , and $f(y_2)$ is the marginal (joint) density of the random vector Y_2 . However, this model restricts Y_1 to be one dimensional under its current formulation. As a possible extension of the proposed univariate model for the semi-continuous variable Y , the model can be represented as $Y_1 = V(Y_2)W$, where W is a binary random variable and $V(Y_2)$ is a continuous positive-valued random variable, whose distribution depends on Y_2 . Finally, each vector in Y_2 assumes a conditionally specified distribution.

Firstly, $Z_{2i} = \log(Y_{2i})$ (where the natural logarithm is taken with respect to each component of the vector Y_{2i}) is parameterized, which will be distributed as a multivariate normal, with mean vector, $x_i B^{(2)}$ and variance matrix, Σ_{22i} , allowing the model to be heteroskedastic. $B^{(2)}$ is $(K \times (J - C))$ and Σ_{22i} is $(J - C) \times (J - C)$, where J is equal to the number of dependent variables in the multivariate model and C is the number of semi-continuous dependent variables each taking the value 0 with positive probability. The assumption of log-normality is often made due to the ease with which a multivariate log-normal can be computed and its ability to account for skewness. This distribution can be expressed as $Z_{2i} \sim N(x_i B^{(2)}, \Sigma_{22i})$, where Z_{2i} is a $((J - C) \times 1)$ -dimensional vector of outcomes. This formulation, similar to the univariate case, allows for the use of separate parameters for the binary variables and positive-valued continuous part in addition to capturing the heteroskedasticity.

The conditional probability of Y_1 given Y_2 is modelled through a mixture of modelling mechanisms

that take into account the realizations from Y_2 such that $Y_{1i}|Y_{2i}=y_{2i} \sim \text{SLNM}(\rho_i, \mu_i(y_{2i}), \sigma_i^2(y_{2i}))$, where SLNM is as defined earlier. More specifically, $\mu_i(y_{2i}) = x_i B^{(1)} + \sum_{12i} \sum_{22i}^{-1} (y_{2i} - x_i B^{(2)})$ and $\sigma_i^2(y_{2i}) = \sum_{11i} - \sum_{12i} \sum_{22i}^{-1} \sum_{21i}$, where $B^{(1)}$ is $(K \times C)$. The covariance matrix Σ_i can be divided into the following components, where Σ_{11i} refers to the covariance matrix of semi-continuous variables, Σ_{22i} refers to the covariance matrix of continuous variables, while Σ_{12i} and Σ_{21i} measure the cross-covariances between the two components, at observation i . (One must be careful in specifying the submatrices within Σ_i to make sure that for any observed value of the predictors X_i 's the full $J \times J$ matrix Σ_i is a positive-definite (p.d.) matrix for each i)

$$\Sigma_i = \begin{pmatrix} \Sigma_{11i} & \Sigma_{12i} \\ \Sigma_{21i} & \Sigma_{22i} \end{pmatrix} \tag{5}$$

When $C=1$, this leads to the following conditional probability density function for Y_{1i} given $Y_{2i}=y_{2i}$:

$$\begin{aligned} f(y_{1i}|y_{2i}) &= 1 - \rho_i(\gamma), & \text{if } y_{1i} = 0 \\ &= \rho_i(\gamma) \frac{1}{y_{1i} \sigma_i(y_{2i})} \phi\left(\frac{\log(y_{1i}) - \mu_i(y_{2i})}{\sqrt{\sigma_i^2(y_{2i})}}\right), & \text{if } y_{1i} > 0 \end{aligned} \tag{6}$$

Ghosh *et al.* (2006) demonstrate through simulation studies that similar models have better finite sample performance with tighter interval estimates when using Bayesian procedures instead of classical maximum likelihood methods. Due to these advantages, the previously developed models will use recently developed Bayesian techniques. In order to develop a Bayesian model, the sampling distribution is weighted by prior distributions. The likelihood function $L(\cdot)$ is proportional to the sampling density $f(\cdot)$, where the proportionality constant is allowed to be a function of the observed data, such that $L(\theta, y) \propto f(y|\theta)$, where θ represents the collection of all parameters, which for the present purposes will include $\theta = (\beta, \alpha, \gamma)$ in the case of a univariate response variable. While prior assumptions can have some effects in small samples, this influence is known to diminish with larger sample sizes (by the use of well-known Bernstein–von Mises theorem). Additionally, prior assumptions can be uninformative in order to minimize any effects in small samples (Kaas & Wasserman 1996). For each parameter in the model, the following vague normal prior is assumed, where $\theta \sim N(0, \Lambda)$ such that $\theta = \{(\beta_{kj}, \alpha_{kj}, \gamma_{kc}) \text{ for } k=1, \dots, K, j=1, \dots, J, \text{ and}$

$c=1, \dots, C\}$ and Λ is a known matrix of dimension $m \times m$, where $m=K(2J+C)$. The given formulation applies to univariate versions when $J=1$ and $C=1$. Additionally, the diagonal elements of Λ can be chosen large enough to make the prior relatively uninformative.

Given the preceding specifications of a sampling density and prior distributions, a full hierarchical Bayesian model can be easily developed in principle. However, due to the difficulty in deriving the posterior distribution of the high-dimensional parameter θ analytically, Markov Chain Monte Carlo (MCMC) methods are used to obtain samples from the posterior distribution of θ using the WinBUGS programming software. MCMC methods allow for the computation of posterior estimates of parameters through the use of MC simulation based on Markov chains from a posterior distribution with the posterior distribution as its stationary distribution. The draws arise from a Markov chain since each draw depends only on the last draw, which satisfies the Markov property. As the posterior density being the stationary distribution of such a chain, the samples obtained from the chain are approximately generated from the posterior distribution following a burn-in of initial draws, i.e. samples are retained after discarding few initial samples with the assumption that the latter samples are drawn from the stationary distribution of the Markov chain.

Bayesian point and interval estimates are computed by using the empirical summaries of the MCMC samples generated from the posterior distribution. In a Bayesian framework, interval estimates are typically computed using the highest posterior density region, which will be different from a classical confidence interval when posterior distributions are bi-modal or asymmetric. As long as posterior estimates do not show bi-modal attributes, the interval between the 2.5 and 97.5 percentiles can be taken to test whether the variable is significantly different from zero. Since multiple chains are usually used in Bayesian analysis, the potential scale reduction factor, or \hat{R} , measure developed in Brooks & Gelman (1998) is popularly used to monitor convergence. This measure essentially uses scaled ratios between pooled and within variances of the multiple chains so that when $\hat{R} \approx 1$, it can be concluded that all chains have mixed well despite different multiple starting values and have been generated from the same target distribution.

It should be noted that the well-known identifiability problem of mixture models due to label switching of the indexes of the mixture components does not occur

for the proposed models as Y is a mixture of a $\{0, 1\}$ valued random variable (W) and a continuous random variable (V). In fact, the parameter vector γ can easily be estimated (and hence identified) using the observed indicators $W_i = I(Y_i > 0)$ and the predictor variables (x_i) using a logistic regression model. Further, the remaining components of vector θ (excluding γ) can be estimated by creating a subset of data $Y_i = V_i W_i$ using only those observation i 's for which $Y_i > 0$. Thus, the parameters γ and remaining components of θ can be estimated using a sequence of logistic model and a likelihood based on V_i 's as defined above.

Predictive values within a Bayesian framework arise from the posterior predictive distributions, which is distinctly different from classical theory that uses a conditional predictive density obtained by plugging in the estimated parameter in the sampling density and thus ignoring the uncertainty of the conditional density. A Bayesian posterior predictive density, however, is obtained by integrating out the sampling density (i.e. the conditional density of data given the parameters) with respect to the posterior density. In the SLNM model, predicted values will be the product of two posterior mean estimates. MCMC methods allow sampling from the posterior density function that combines both the prior distributional information and the sampling distribution (likelihood function). Formally, prediction in the SLNM model is characterized by $\hat{y}_i = v_i w_i$, where v_i and w_i are generated from their predictive distributions. $\log(v_i)$ is from a normal distribution with mean ($\mu_i = x_i' \hat{\beta}$) and variance ($\sigma_i^2 = \exp(x_i' \hat{\alpha})$), while w_i is from a Bernoulli distribution with parameter $\rho_i(\hat{\gamma})$, where $\hat{\beta}$, $\hat{\alpha}$ and $\hat{\gamma}$ are MCMC draws from their posterior distributions, where x_i represents a vector of predictor variables at which prediction values are desired. Since many draws from a Bernoulli distribution will result in 0 and 1 outcomes, the mean will lie between zero and unity. To allow for prediction of both zero and positive values, the median of the Bernoulli draws was used for prediction. This allows for observations to fully take on the continuous random variable if more than half of the time it was modelled to do so, while those that are more likely to take on zero values, as indicated by the Bernoulli outcomes, take on a zero value.

In order to assess the predictive performance of the model, a randomly selected two-thirds of the observed data set is used for estimation, while the remaining one-third is used for prediction, allowing both model fit measures and out-of-sample predictive power performance to be tested. In the present study,

estimations based on a Tobit regression also incorporate Bayesian estimation techniques. To compare in-sample measures of model fit, the deviance information criteria (DIC) (Spiegelhalter *et al.* 2002), which is similar to its frequentist counterpart, Akaike's Information Criteria (Akaike 1974) in which smaller values indicate improved model fit, is used.

To measure the out-of-sample predictive power within a modelling strategy, the mean squared prediction error (MSPE) associated with the final third of randomly selected data is computed. MSPE is formulated as $MSPE = (1/M) \sum_{i=1}^M (\hat{y}_i - y_i)^2$, where M is some proportion of the full data sample, such that $M = [n/b]$. For the present purposes $b = 3$, which allows for prediction on the remaining third, based on estimates from the first two-thirds. This allows for a sufficient amount of observations available for estimation and prediction. MSPE is then decomposed into error due to central tendency (ECT), error due to regression (ER) and error due to disturbances (ED), based on Bibby & Toutenburg (1977). The parameters are estimated based on the simulated dataset, given the above specifications. MCMC sampling is used for Bayesian estimation with a burn-in of 1000 iterations for the SLNM model and 20000 for the Tobit model. After discarding the burn-in samples, another 10000 iterations are used with a thin rate of 15 for univariate estimation with three parallel Markov chains. Multivariate simulations make use of 6000 samples on a single Markov chain. The numbers of burn-in and sample iterations were selected based on closely monitoring \hat{R} , based on Brooks & Gelman (1998). The burn-in rate was substantially higher with the Tobit model than the SLNM model, as a few parameter estimates took particularly long to converge.

A SIMULATION STUDY COMPARING MODELLING PERFORMANCES

This section focuses on simulating data in order to compare the goodness-of-fit for the two previously specified models. The major advantage of evaluating model performances using a simulated set 245 of data is that the true form of the data generation process is known, which is generally not possible with real data sets. Various goodness-of-fit criteria are evaluated to compare the performance of a given models when the data are generated either from the same model (e.g. Tobit) or from another model (e.g. SLNM). These simulated data scenarios allow evaluation of the extent

Table 1. Model comparison based on data simulated from univariate Tobit and SLNM models ($n=10$)

n	Est. model	Data simulated from			
		Tobit model		SLNM model	
		MSPE	DIC	MSPE	DIC
200	Tobit	0.34 (0.024)	214 (6.1)	9 (4.8)	267 (20.9)
	SLNM	0.44 (0.070)	230 (5.2)	8 (4.7)	105 (12.8)
500	Tobit	0.29 (0.015)	512 (9.5)	5 (2.7)	747 (43.9)
	SLNM	0.35 (0.020)	586 (9.4)	5 (2.6)	244 (21.8)
1000	Tobit	0.28 (0.010)	1000 (10.0)	3 (1.4)	200 (385.7)
	SLNM	0.32 (0.013)	1148 (16.4)	3 (1.4)	354 (31.3)

Note: MCSEs are reported in parentheses.

of bias induced by fitting a model to a data set that is generated from possibly a different model.

In simulating data, there are three key characteristics that align the simulated data set with the cattle production data set to be used in the empirical study, which include heteroskedastic errors (Belasco *et al.* 2009a), a degree of censoring around 0.50 and positive skewness. Data are repeatedly simulated based on samples sizes of 200, 500 and 1000. The first simulated model will be a heteroskedastic Tobit model, which fundamentally assumes a normally distributed latent variable, y_i^* such that, $y_i^* = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$ and the dependent variable $y_i = \max(y_i^*, 0)$ can be observed. In order to account for heteroskedasticity, $\varepsilon_i \sim N(0, \sigma_i^2)$ is assumed, where $\sigma_i^2 = \exp(\alpha_0 + \alpha_1 x_{1i})$.

In this scenario, the censored and uncensored variables (or continuous and discrete portions of the semi-continuous distribution) come from the same data generation process. Simulated values are based on $(\beta_0, \beta_1) = (-3.0, 1.0)$ and $(\alpha_0, \alpha_1) = (2.0, 0.2)$ where starting seeds are set in order to replicate results. For each sample size, values for x_i are generated randomly based on a uniform distribution bounded at 1 and 5. For the above set of selected values of the parameters, the percentage of observed zeros these simulated data sets ranged from 0.45 to 0.55. In order to compute the sampling variation of the estimates over 10 repetitions, Monte Carlo standard errors (MCSEs) are reported. More precisely, the sampling variation of an estimator $\hat{\theta}$ can be measured by $MCSE_{\hat{\theta}} = \sqrt{[1/N(N-1)] \sum_{j=1}^N (\hat{\theta}_j - \hat{\theta})^2}$, where $\hat{\theta}_j$ is the estimate obtained from the j th repetition and N denotes the number of MC repetitions. For all data generation scenarios, we performed $N=10$ MC repetitions.

As an alternative to the preceding simulation scenario, data were also generated from an SLNM model. The major distinctions between this simulation and the previously developed Tobit-based data set are that the probability of a censored outcome is modelled separately and that the latent variable is positively skewed. For this SLNM model, the following values of the parameters were selected: $(\beta_0, \beta_1) = (-1.5, 0.5)$, $(\gamma_0, \gamma_1) = (-5.0, 1.7)$ and $(\alpha_0, \alpha_1) = (2.0, 0.2)$. The proportion of observed zeros in this second scenario also ranged from 0.45 to 0.55, similar to the first scenario.

Results from simulations based on a Tobit regression and an SLNM regression are reported in Table 1. As expected, the performance of a correctly specified model is better than that of the wrong model. Lower MSPE indicates that the prediction of the out-of-sample portion of the data set favours that of the true model at all sample levels. MSPE penalizes observations with large residuals that tend to be more prevalent as the dependent variable value increases. Gurmu & Trivedi (1996) point out that mixture models tend to over-fit data. By over-fitting the data, model fit tests might improve while prediction remains less accurate. The Tobit model could outperform the SLNM due to the lack of positive skewness or the loss in degrees of freedom from irrelevant variables as part of the two-step process. However, a semi-continuous normal mixture model was also used and while the performance was superior to the SLNM, it still did not outperform the Tobit.

When data are generated from an SLNM model, the results indicate a superior model fit with the assumed model, relative to the Tobit model and a substantially lower MSPE, indicating better out-of-sample prediction performance, partially due to its ability to account for positive skewness. When the sample size is $n=500$,

Table 2. Model comparison based on data simulated from multivariate Tobit and SLNM models ($n = 10$)

n	Est. model	Data simulated from					
		Tobit model			SLNM model		
		MSPE		DIC	MSPE		DIC
Y_1	Y_2	Y_1	Y_2				
200	Tobit	0.75 (0.067)	0.27 (0.044)	554 (11.7)	0.99 (0.081)	0.82 (0.262)	577 (10.5)
	SLNM	0.75 (0.068)	0.44 (0.066)	600 (10.4)	1.00 (0.083)	0.56 (0.226)	422 (9.0)
500	Tobit	0.97 (0.089)	0.29 (0.031)	1416 (30.1)	1.00 (0.026)	0.66 (0.036)	1481 (24.5)
	SLNM	0.97 (0.090)	0.34 (0.030)	1499 (27.2)	1.00 (0.027)	0.22 (0.030)	1051 (15.1)
1000	Tobit	0.87 (0.034)	0.28 (0.022)	2826 (24.5)	0.96 (0.031)	1.10 (0.203)	3022 (61.7)
	SLNM	0.87 (0.034)	0.34 (0.020)	3013 (22.2)	0.96 (0.030)	0.56 (0.138)	2056 (20.7)

Note: MCSEs are reported in parentheses.

the relative loss (in terms of having lower DIC) by fitting an SLNM to data sets generated by a Tobit model is $c. 0.14$ ($(586.331 - 512.361)/512.361$) only, whereas the relative loss of fitting a Tobit model to data sets generated from an SLNM is 2.06 ($(747.034 - 244.106)/244.106$). However, the relative loss (in terms of having lower MSPE) was about the same for each incorrectly assumed model. Thus, although by fitting an SLNM to Tobit-generated data sets DIC might be reduced there is a much larger loss in model fit when a Tobit is incorrectly assumed.

While univariate censored data are widely used in empirical studies, multivariate censored data are not as widely utilized, despite having many applications in production and agricultural economics. For this reason, it is important to simulate a multivariate data set that comes from both a Tobit and a multivariate version of the SLNM model. The Tobit model is constructed by first generating latent vectors (Y_{1i}^* , Y_{2i}^*) from a bivariate normal distribution with mean vector ($\beta_{11} + \beta_{12x_i}$, $\beta_{21} + \beta_{22x_i}$) and logarithm of the variance vector ($\alpha_{11} + \alpha_{12x_i}$, $\alpha_{21} + \alpha_{22x_i}$) and a fixed covariance of 0.6 for all $i = 1, \dots, n$. The dependent variables are generated by setting $Y_{1i} = Y_{1i}^*$ for all i and $Y_{2i} = \max(0, Y_{2i}^*)$, and thus only Y_2 values are censored at zero whereas Y_1 has a continuous distribution. The true value of the parameters were set at $\beta = (0.0, 4.0; -1.5, 0.5)$, $\alpha = (1.5, 0.5; 2.0, 0.2)$, and covariance off-diagonal element, $t_{12} = 0.6$. As in the univariate case, a second scenario was also generated which consisted of generating data from the multivariate extension of the SLNM, and the same parameter values were used as in the bivariate Tobit model with the additional parameter set at $\gamma = (-5.0, 1.7)$. As in the bivariate

Tobit model, Y_1 was set as a fully continuous variable, while Y_2 was semi-continuous with censoring in nearly half of its observations.

The results based on generating data from a bivariate Tobit model are shown in Table 2. It is not surprising to note that correct model specification matters in terms of within-sample fit, as DIC is lower when the assumed model matches the simulated model. The SLNM demonstrates an improved advantage (in terms of lower MSPE) when using the multivariate model and the assumed model is correct. This is probably because the ability to predict Y_2 is improved substantially in the SLNM from the use of a conditional distribution. Given the present hypothesis that cattle MORT are more accurately characterized from an SLNM model, it might be assumed that within-sample fit might be stronger than with the Tobit, while prediction improves marginally.

AN APPLICATION TO CATTLE PRODUCTION YIELDS

This section applies the preceding models to cattle production risk variables. The proposed SLNM regression model is hypothesized to characterize semi-continuous cattle MORT better than the Tobit model because of the flexibility of modelling the two components of the MORT distribution using a separate set of parameters. In the previous section, repeated simulations were used to mimic this type of data, which is characterized by heteroskedastic errors, a high degree of censoring and a system with cross-equation correlation. Results from such MC simulations demonstrate the power of the proposed SLNM

model to fit data generated based on itself and the Tobit model as well, in both univariate and multivariate situations.

Past research has focused on modelling agricultural yields, while research dealing specifically with semi-continuous valued yields is limited. The main reason for the lack of research into censored yields is because crop yields are not typically censored at upper or lower bounds. However, with the emergence of new live-stock insurance products, new yield measures must be quantified in order for risks to be properly identified. In contrast to crop yield densities, yield measures for cattle health possess positive skewness, such as the MORT and veterinary costs. Crop yield densities typically possess a degree of negative skewness as plants are biologically limited upward by a maximum yield, but can be negatively influenced by adverse weather, such as drought. Variables such as mortality have a lower limit of zero, but can rise quickly in times of adverse weather, such as prolonged winter storms or disease.

Cattle MORT are hypothesized to follow a two-step process because pens tend to come from the same, or nearby, producers and are relatively homogeneous. Therefore, a single mortality can be seen as a sign that a pen is more prone to sickness or disease and may distinguish a 'healthy' pen from an 'unhealthy' pen. Additionally, airborne illnesses are contagious and can be spread rather quickly throughout the pen. Other variables that describe cattle production performance are introduced and evaluated using the previously developed multivariate framework. These variables include dry matter feed conversion (DMFC), which is measured as the average kg of feed a pen of cattle require to add 1 kg of weight gain; average daily gain (ADG), which is the average daily weight gain (kg) per head of cattle; and veterinary costs per head (VCPH), which is the amount of VCPH that are incurred over the feedlot stay.

Cattle MORT from commercial feedlots are of particular interest due to their importance in cattle feeding profits. Typically, MORT are zero or small, but can rise significantly during adverse weather, illness or disease. The data used in the present study consist of five commercial feedlots residing in Kansas and Nebraska, and include entry and exit characteristics of 11 397 pens of cattle at these feedlots from 1996 to 2005. Of the five feedlots, two are closely located to one another in Kansas, while the remaining are located in close proximity to each other within Nebraska. Preliminary analysis suggested the use of aggregating

feedlots into two clusters that are distinguished by state. The data did not exhibit any time trends or structural breaks in dependent variables (ADG, DMFC, MORT and VCPH) over the 10 years. Independent variables are also included that are hypothesized to impact on the mean and variance of the dependent variables. Independent variables included in the present study include entry weight (InWt) and categorical variables to indicate gender (steers, heifers and mixed); whether the pen was placed in a Kansas feedlot (Kansas) as opposed to a feedlot in Nebraska; and season of placement in order to indicate whether the pen was placed on a feedlot between December–February (Winter), March–May (Spring), June–August (Summer) or September–November (Autumn). [Table 3](#) presents a summary of characteristics for different levels of MORT, including no mortalities.

Particular attention will be placed on whether zero or positive MORT can be strongly determined based on the data at hand. The degree of censoring in this sample is 0.46, implying that almost half of the observations contain no mortality losses. There is strong evidence that MORT are related to the previously mentioned conditioning variables, but whether censored mortality observations are systematically different from observed positive values must be determined. Positive MORT may be a sign of poor genetics coming from a particular breeder or sickness picked up within the herd. The assumption is that the cattle within the pen are quite homogeneous. Homogeneity within the herd is desirable as it allows for easier transport, uniform feeding rations, medical attention and the amount of time on feed. If homogeneity within the herd holds, then pens that have mortalities can be defined as a distinct type that is separate from those with no mortalities.

However, mortalities may also occur without warning and for unknown reasons. Glock & DeGroot (1998) report that 0.40 of all cattle mortalities in a Nebraska feedlot study were directly caused by sudden death syndrome. Glock & DeGroot (1998) loosely define sudden death as any case where feedlot cattle are found dead unexpectedly, but they also point out that these deaths were without warning, which could be due to a 'sudden death' or lack of observation by the feedlot workers. Smith (1998) also reports that respiratory disease and digestive disorders are responsible for c. 0.44 and 0.25 of all mortalities, respectively. The high degree of correlation between dependent variables certainly indicates that lower MORT can be associated with different performance in

Table 3. Mean comparison of pens using average values of both dependent and predictor variables with differing mortality losses

Variable	Units	MORT (%)*					
		0	0.01–1	1–2	2–3	3–4	>4
No. of observations		5161	2415	2327	744	305	445
DMFC	kg (gain/feed)	6.05 (0.63)	6.27 (0.65)	6.21 (0.67)	6.34 (0.72)	6.42 (0.69)	6.85 (1.38)
ADG	kg/head day	1.59 (0.21)	1.52 (0.20)	1.48 (0.20)	1.41 (0.19)	1.39 (0.17)	1.28 (0.21)
VCPH	US\$/ head	10.18 (3.79)	10.08 (3.81)	12.46 (5.79)	15.67 (7.09)	17.89 (8.40)	26.57 (11.45)
InWt	kg	342.13 (36.61)	340.89 (37.84)	326.46 (38.89)	311.32 (38.94)	317.06 (42.32)	304.98 (43.73)
Proportion of sample							
Steers		0.53	0.56	0.49	0.43	0.42	0.35
Mixed		0.10	0.07	0.14	0.20	0.20	0.32
Heifers		0.36	0.37	0.37	0.36	0.38	0.33
Kansas		0.82	0.76	0.82	0.77	0.79	0.82
Nebraska		0.18	0.24	0.18	0.23	0.21	0.18
Winter		0.24	0.27	0.26	0.29	0.29	0.21
Autumn		0.23	0.23	0.27	0.30	0.30	0.39
Spring		0.26	0.24	0.21	0.16	0.17	0.10
Summer		0.27	0.26	0.26	0.25	0.24	0.30

* Note: A MORT that results in a whole number is placed into the higher bins (i.e. 3.00% is placed in 3–4 bin). Standard errors are in parentheses.

DMFC, dry matter feed conversion; ADG, average daily gain; VCPH, veterinary costs per head.

the pen. However, the question in the present study will be whether positive MORT significantly alter the performance. For this reason, additional parameters were estimated to examine the likelihood of a positive mortality outcome in the mixture regression model.

A recent study by Belasco *et al.* (2009b) found that the mean and variance of MORT in cattle feedlots are influenced by entry-level characteristics such as location of the feedlot, placement weight, season of placement and gender, all of which are used in the present study. The influence of these parameters will be estimated using the previously formulated models, based on two-thirds of the randomly selected data set where $n = 7598$. The remaining portion of the data set, $m = 3799$, will be used to test out-of-sample prediction accuracy. Predictive accuracy is important in existing crop insurance programmes, where past performance is used to derive predictive density functions for current contracts.

ESTIMATION RESULTS

The results from using a univariate Bayesian Tobit model with heteroskedastic errors to model cattle MORT are shown in Table 4. Tobit estimates for β measure the marginal impact of changes in the conditional variables on the latent MORT (Y^*). To measure the approximate marginal impact on the observed

MORT (Y), the following relationship from Greene (2008) is used, where $dE[y_i|x_{ki}]/dx_{ki} = \beta_k \times \Pr(y_i^* > 0)$. For example, the coefficient corresponding to in-weight states that a one-unit increase in logarithm of the entry weight decreases the observed MORT by 1.7% ($= 3.9\% \times 0.54$). The estimates for α measure the relative impact on the variance, since $\alpha_k = (d\sigma_i^2/dx_k) / (1/\sigma_i^2)$. For example, the estimated coefficient corresponding to autumn implies that a pen placed in the autumn months is associated with a variance that is 40.7% lower than pens placed in summer months.

Next, the same set of data is estimated using the previously developed SLNM model in order to test the hypothesis that they will have a better fit. Results from the SLNM model are also shown in Table 4. Parameter estimates in the SLNM model refer to two distinct processes. The first process includes the likelihood of a zero outcome or one described by a log-normal distribution. Based on this formulation, the parameter estimates for γ can be expressed as the negative of the marginal impact of the conditional variable k on the probability for individual i of a positive outcome, relative to the variance of the Bernoulli component:

$$\begin{aligned} \gamma_k &= -\frac{\partial \rho_i}{\partial x_{ki}} \cdot \frac{1}{\rho_i} \cdot \left[\frac{1 + \exp(x'_i \gamma)}{\exp(x'_i \gamma)} \right] \\ &= -\frac{\partial \rho_i}{\partial x_{ki}} \cdot \frac{1}{\rho_i(1 - \rho_i)} \end{aligned} \tag{7}$$

Table 4. Estimates of fed cattle mortality parameters based on a univariate Tobit and SLNM models

Variables	Tobit			SLNM		
	Est.	P	s.d.*	Est.	P	s.d.
Mean effects (β)						
Intercept	25.9	<0.05	1.04	9.5	<0.05	0.76
Steers	0.22	<0.05	0.061	0.04	NS	0.024
Mixed	0.47	<0.05	0.125	0.38	<0.05	0.036
Kansas	-0.07	NS	0.064	0.13	<0.05	0.026
Log (lnWt)	-3.9	<0.05	0.16	-1.4	<0.05	0.12
Winter	0.03	NS	0.073	-0.03	NS	0.028
Autumn	0.03	NS	0.081	0.06	NS	0.030
Spring	-0.12	NS	0.072	-0.09	<0.05	0.029
Variance effects (α)						
Intercept	-10.5	<0.05	0.97	-2.4	<0.05	0.81
Steers	0.12	<0.05	0.053	0.12	<0.05	0.049
Mixed	-0.82	<0.05	0.075	-0.20	<0.05	0.069
Kansas	-0.23	<0.05	0.060	0.10	NS	0.053
Log (lnWt)	1.41	<0.05	0.15	0.47	<0.05	0.13
Winter	0.02	NS	0.065	0.06	NS	0.062
Autumn	-0.41	<0.05	0.066	-0.14	<0.05	0.062
Spring	0.32	<0.05	0.070	0.24	<0.05	0.064
Probability limit effect (γ)						
Intercept				-19.3	<0.05	1.78
Steers				-0.14	<0.05	0.056
Mixed				-0.29	<0.05	0.079
Kansas				0.10	NS	0.061
Log (lnWt)				2.9	<0.05	0.27
Winter				-0.07	NS	0.065
Autumn				-0.09	NS	0.065
Spring				0.17	<0.05	0.066
DIC	23 020			20 752		
MSPE†	2.461			2.612		
ECT	0.350		(0.014)‡	0.084		(0.032)‡
ER	1.204		(0.489)‡	0.316		(0.121)‡
ED	0.907		(0.368)‡	2.212		(0.847)‡

* Standard deviations (s.d.) are reported for continuous variables, while standard error differences (s.e.d.) are used for all categorical variables.

† MSPE is decomposed into ECT, ER and ED (Bibby & Toutenburg 1977).

‡ Proportion of error.

NS: not significant at $P < 0.05$.

where the variance is shown as $\rho_i(1 - \rho_i)$. For example, entry weight largely and negatively influences the likelihood of positive MORT, as indicated by the positive parameter estimate ($\gamma_5 = 2.901$). This is not surprising given that more mature pens are better equipped to survive adverse conditions, whereas younger pens tend to be more likely to result in mortalities. Alternatively, mixed pens (mixed) have a negative γ coefficient ($\gamma_3 = -0.289$) which implies that there is a positive relationship, relative to heifer pens. Estimates for steers and mixed indicate the difference

in probability of a positive mortality, relative to a heifer pen. Therefore, if a pen is mixed, it has a higher probability of incurring positive mortality realizations that can be modelled with a log-normal distribution.

In the SLNM results, when γ coefficients are negatively related to β coefficients it implies that covariates impact on the mean of positive and observed values in the same direction as the likelihood of a positive value. For example, increases to entry weight shift the mean of MORT downward ($\beta_5 = -1.434$) and also decrease the probability of a positive

Table 5. *Out-of-sample prediction results for Tobit and SLNM models*

	Proportion correctly predicted*		Kendell's Tau	MSPE			
	0	>0		0	0–1	1–2	>2
Observed	1740	2059		1741	795	797	466
Tobit	0.46	0.67	0.14	0.18	0.32	1.26	16.68
SLNM	0.39	0.72	0.11	1.98	0.94	0.90	10.77

* Note: Correct prediction corresponds to predicting a zero or positive value that matches the actual value.

outcome ($\gamma_5 = 2.901$). This does not necessarily mean that the two processes work identically, as is assumed with the Tobit model, but rather tend to generally work in the same direction.

Parameter estimates for β refer to the marginal impact that the conditioning variables have on the positive realizations of MORT. Interpretations for these parameters refer to the marginal increase in the log of MORT. For example, a one-unit increase in the logarithm of entry weight is associated with a reduction in MORT by 1.4% ($\beta_5 = -1.434$) for the observations that experience a positive MORT.

It is interesting to note the different implications from parameter estimates from the Tobit and SLNM models. For example, an insignificant mean difference parameter estimate for the variable Kansas in the Tobit model implies that MORT are not significantly impacted by feedlot location. However, parameter estimates from the SLNM model infer that pens placed into feedlots located in Kansas have a lower likelihood of a positive MORT ($\gamma_4 = 0.099$), relative to Nebraska feedlots. At the same time, pens placed in Kansas that have a positive MORT can be expected to realize logged MORT that is 0.13 higher than Nebraska feedlots ($\beta_4 = 0.130$). It might be strange to have significant impacts in opposite directions that influence both the likelihood of mortality and the positive MORT, but by distinguishing between these processes their respective impacts can be isolated. One possible explanation might be that Kansas lots spend more time on preventing mortalities from occurring through vaccinations or backgrounding, but are not able to detect the spread of disease as quickly as the Nebraska feedlots. This is a notable departure from the Tobit model, which saw no significant influence since these impacts essentially cancelled each other out.

The SLNM model also demonstrates a better overall ability to fit the within-sample data with a DIC measure of 20 752, as compared to the Tobit model, which has a DIC value of 23 020. However, MSPE is minimized

when using the Tobit model, indicating a trade-off between within sample fit and out-of-sample prediction effectiveness.

Table 5 breaks MSPE into four distinct groups based on actual values. The SLNM model does not appear to do as well fitting low actual values, relative to the Tobit. The SLNM model appeared to do much better in correctly predicting the positive actual values, as only 0.39 of the actual zero values were correctly predicted, while 0.72 of the positive values were correctly predicted to be positive values. A lower Kendall's Tau for the SLNM model also demonstrates that the Tobit more accurately predicts whether a value is zero or positive. This is likely due to the inability of the given *ex ante* variables to explain much of the variation in the model which is also shown in past studies (Belasco *et al.* 2009b). While the results shown in Table 4 reflect significant estimates corresponding to γ , they may not be large enough in magnitude to have a real impact on the probability of positive observations. To illustrate, the mean of predicted values for ρ_i was 0.544, which is close to the actual degree of censoring in the out-of-sample values (0.55). However, the standard deviation around predicted ρ_i was 0.07. The implications of this lack of spread means that the model will tend to over-predict the amount of non-censored values. The SLNM model does do very well in fitting actual positive values that are greater than one, which is likely due to its accommodation for positive skewness.

Regressions from univariate mortality models offer information concerning the relative impact each conditioning variable has on MORT. However, this variable is likely better characterized in a multivariate setting with other variables that explain the health and performance of cattle on feedlots, ultimately describing production risk in cattle feeding enterprises. To this end, the multivariate Tobit model and multivariate SLNM models were used to characterize these four variables, described earlier.

Table 6. *Model fit and predictive power estimates of fed cattle parameters based on a multivariate Tobit and SLNM models*

Model	MSPE				DIC
	DMFC	VCPH	ADG	MORT	
Tobit	0.008	0.182	0.169	1.528	23 284
SLNM	0.008	0.182	0.170	1.296	19 348

The multivariate SLNM model is applied to the cattle feedlot data set and more accurately fits the within-sample data and predicts with more accuracy, relative to the Tobit model, as shown in Table 6. DIC for the SLNM model is substantially lower than with the Tobit model, mainly due to the more accurate fit for MORT, which contributes quite a lot of unexplained variability to the system of equations. MSPE is significantly lower for semi-continuous variables when applied to the multivariate mixture model. This is particularly interesting given the contradictory univariate results; however, these results are more consistent with data generated from an SLNM model in the simulation study. MORT are shown to be more accurately characterized in a multivariate setting because of the effects from other non-censored variables as MSPE improves significantly from 2.612 (shown in Table 4) to 1.296 (shown in Table 6). Recall, that the expected value and variance of MORT accounts for the uncensored variable levels in the multivariate setting. This link allows for information from the continuous variable to feed into the semi-continuous variable on non-continuous portions of the density.

DISCUSSION

The present paper provides two distinct contributions to existing research. The first is to develop an alternative modelling strategy to the Tobit model and more traditional mixture models, specifically for semi-continuous data. This model originates in a univariate case similar to that of Duan *et al.* (1983), then is extended to allow for multivariate settings. In the multivariate setting, the covariate relationship accounts for the correlation between semi-continuous and continuous variables. The second contribution is to more accurately describe production risk for cattle feeders by examining the model performance of different regression techniques. MORT play a vital role in cattle feeding profits, particularly due to the skewed

nature of this variable. A clearer understanding of mortality occurrences will assist producers as well as private insurance companies, who offer mortality insurance, in managing risk in cattle operations. Additionally, production risk in cattle feeding enterprises play a significant role in profit variability, but is currently uninsured by current federal livestock insurance programmes. Existing livestock insurance programmes include the Livestock Revenue Product, which provides insurance against drops in cattle output prices, and the Livestock Gross Margin, which insures against the net impact of cattle output price drops and corn input price spikes. An accurate characterization of production risk plays an important role in addressing risk for producers and insurers.

Modelling semi-continuous data remains a large area of concern and current research in econometrics. While use of the Tobit model may be well-justified in certain instances, the results from both simulated and actual cattle production data sets suggest the use of an SLNM regression model. This is particularly true in instances where data come from a two-step process. While two-step processes have been applied to Hurdle models, mixture regression models have largely been ignored in economic studies. This is mainly a result of the past limitation of mixture regression models to count data (zero-inflated models). In the present study, a mixture regression model is developed that can handle both univariate and multivariate situations efficiently, in addition to nesting the standard Tobit model. Additionally, the inherent parametric flexibility allows for distributional assumptions to change based on the data on hand, rather than strictly using truncated or normal distributions. Here, a log-normal distribution is used to capture the positively skewed nature of cattle MORT, which gives the SLNM model significant advantages over the Tobit model. Advantages in model fit for the SLNM model stem from the ability to isolate and identify the impacts from observing a positive MORT and the level of MORT.

Production risk in cattle feeding enterprises is inherently complex, given the many areas risk can originate. The results from the present research demonstrate the potential gains from using this particular mixture model. Before applying this model to the data, simulations were conducted to test the model's ability to predict and fit data generated in different forms. These simulations provide results that support the use of the mixture model, in both prediction and model fit, when the data are from a two-step process. However, the mixture model did not

demonstrate a strong ability to fit the data when the data are generated based on a Tobit model.

A solid understanding of cattle production risk is limited by the ability to characterize variability. The proposed model takes a step forward in developing a modelling strategy that can be used to measure other livestock or live animal productive measures. By more accurately characterizing these risks insurance companies, animal producers and operators can better understand the risks involved with animal production. Additionally, a more general form allows for a more accurate depiction of *ex ante* variables and their impact on MORT and other production variables. Future research is currently focusing on developing this model to account for systems where censoring occurs in more than three variables, which is currently problematic in classical estimation techniques. Examples where this particular model might be useful include health outcomes, consumption, livestock disease spread or production processes. The framework detailed in the present paper naturally extends to situations of more complex censoring systems.

The major flexibility in the proposed model lies in the ability to make different distributional assumptions. Distributional assumptions typically need to be made in cases when data cannot fully explain variability. However, non-parametric and semi-parametric methods may be of particular interest when large data sets are evaluated, since they allow empirical data to create a unique density. Semi-continuous production and demand systems where each component may not be accurately described by a single distribution may provide an excellent application for semi- or non-parametric extensions of the derived model.

REFERENCES

- AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723.
- AMEMIYA, T. (1974). Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica* **42**, 999–1012.
- AMEMIYA, T. (1984). Tobit models: a survey. *Journal of Econometrics* **24**, 3–61.
- ARABMAZAR, A. & SCHMIDT, P. (1982). An investigation of the robustness of the tobit estimator to non-normality. *Econometrica* **50**, 1055–1064.
- BELASCO, E. J., GHOSH, S. K. & GOODWIN, B. K. (2009a). A multivariate evaluation of *ex ante* risks associated with fed cattle production. *American Journal of Agricultural Economics* **91**, 431–443.
- BELASCO, E. J., TAYLOR, M. R., GOODWIN, B. K. & SCHROEDER, T. C. (2009b). Probabilistic models of yield, price, and revenue risk for fed cattle production. *Journal of Agricultural and Applied Economics* **41**, 91–105.
- BERA, A. K., JARQUE, C. M. & LEE, L. F. (1984). Testing the normality assumption in limited dependent variable models. *International Economic Review* **25**, 563–578.
- BIBBY, J. & TOUTENBURG, H. (1977). *Prediction and Improved Estimation in Linear Models*. London: John Wiley and Sons.
- BROOKS, S. & GELMAN, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**, 434–455.
- CASELLA, G. & BERGER, R. (2001). *Statistical Inference*, 2nd edn. Andover, UK: Thompson Learning Inc.
- CHAVAS, J. P. & KIM, K. (2004). A heteroskedastic multivariate Tobit analysis of price dynamics in the presence of price floors. *American Journal of Agricultural Economics* **86**, 576–593.
- CORNICK, J., COX, T. L. & GOULD, B. W. (1994). Fluid milk purchase: a multivariate Tobit analysis. *American Journal of Agricultural Economics* **76**, 74–82.
- CRAGG, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* **39**, 829–844.
- DUAN, N., MANNING, W. G., MORRIS, C. N. & NEWHOUSE, J. P. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business and Economic Statistics* **1**, 115–126.
- EISWERTH, M. E. & SHONKWILER, J. S. (2006). Examining post-wildfire reseeding on arid rangeland: a multivariate Tobit modelling approach. *Ecological Modelling* **192**, 286–298.
- GHOSH, S. K., MUKHOPADHYAY, P. & LU, J. C. (2006). Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference* **136**, 1360–1375.
- GLOCK, R. D. & DEGROOT, B. D. (1998). Sudden death of feedlot cattle. *Journal of Animal Science* **76**, 315–319.
- GREENE, W. (2008). *Econometric Analysis*, 6th edn, London: Prentice Hall.
- GURMU, S. & TRIVEDI, P. K. (1996). Excess zeros in count models for recreation trips. *Journal of Business and Economic Statistics* **14**, 469–477.
- JONES, A. M. (1989). A double-hurdle model of cigarette consumption. *Journal of Applied Econometrics* **4**, 23–39.
- KAAS, R. & WASSERMAN, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association* **91**, 1343–1370.
- LEE, L. F. (1993). Multivariate Tobit models in econometrics. In *Handbook of Statistics*, Vol. 11 (Eds G. S. Maddala, C. R. Rao & H. D. Vinod), pp. 145–173. Amsterdam: Elsevier.
- SHONKWILER, J. S. & YEN, S. T. (1999). Two-step estimation of a censored system of equations. *American Journal of Agricultural Economics* **81**, 972–982.
- SMITH, R. A. (1998). Impact of disease on feedlot performance: a review. *Journal of Animal Science* **76**, 272–274.
- SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. & VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* **64**, 583–639.
- TOBIN, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* **26**, 24–36.