Feedforward Backpropagation Neural Networks in Prediction of Farmer Risk Preferences

Terry L. Kastens and Allen M. Featherstone

An out-of-sample prediction of Kansas farmers' responses to five surveyed questions involving risk is used to compare ordered multinomial logistic regression models with feedforward backpropagation neural network models. Although the logistic models often predict more accurately than the neural network models in a mean-squared error sense, the neural network models are shown to be more accommodating of loss functions associated with a desire to predict certain combinations of categorical responses more accurately than others.

Key words: neural networks, ordered multinomial logit, risk response prediction.

"The approach to risky choice...is based on the decision maker's personal strengths of belief about the occurrence of uncertain events and his personal evaluation of potential consequences." These words in the preface of Anderson, Dillon, and Hardaker's 1977 decision analysis text highlight the distinctively subjective, personally assessed nature of risk in agriculture. Goodwin and Kastens demonstrate that how farmers feel about risk is correlated with a farmer's crop insurance purchase decision. Subjective assessments of risk are likely to be related to other risk management products as well. However, it is costly to determine an individual's perceptions to risk. If a farmer's feelings toward risk can be accurately predicted using less expensive objective data, firms can segregate classes of potential customers to more efficiently design sales strategies for risk management products.

This study focuses on the question, Can a decision maker's responses to subjective questions regarding agricultural risk be predicted if relevant characteristics of that decision maker are known? Specifically, we examine the accuracy of ordered multinomial logit models compared to feedforward backpropagation neural network models in predicting Kansas farmers' out-of-sample survey responses to risk-related questions. Alternative prediction accuracy test statistics are considered, which may be relevant for different loss functions.

Anderson, Dillon, and Hardaker, and others (Thomas; Schurle and Tierney) suggest that one way to learn how farmers feel about risk is simply to ask them. Risk questions typically solicit a respondent's ranked perception (e.g., from strongly agree to strongly disagree, or on a scale of 1–10). In analysis of such ranked responses, the ordered multinomial logistic regression model (OMLR) has often been the empirical model of choice (Barkley and Flinchbaugh; Vandeveer and Loehman; Kastens and Goodwin). For applied economists, the OMLR model is particularly appealing. Unlike binary or unordered multinomial logit models, an OMLR explicitly takes account of the ranked nature of the choices involved. Because of this more comprehensive use of available information, an OMLR is expected to outperform other models that merely classify responses in unranked classes.

Recently, a new class of empirical models, neural networks, has emerged. A popular subclass of neural networks is feedforward backpropagation neural networks (BPNs). In agricultural economics, BPNs have been ap-

Terry Kastens is assistant professor and Allen Featherstone is associate professor, both in the Department of Agricultural Economics, Kansas State University.

Helpful comments of two anonymous reviewers are gratefully acknowledged. This is Contribution No. 96-176-J from the Kansas Agricultural Experiment Station.

Amer. J. Agr. Econ. 78 (May 1996): 400–415

plied to price forecasting (Grudnitski and Osburn), loan classification (Miller), and production function estimation (Joerding, Li, and Young).

Because BPNs are flexible functional forms, and because in-sample forecasting errors can be driven arbitrarily close to zero, BPNs are routinely validated based on out-of-sample prediction. This has not been the case for conventional econometric models such as OMLR. None of the OMLR-using analyses of survey responses reported earlier used out-of-sample prediction. This is unfortunate in that the research provides little guidance on how an OMLR might perform in predicting survey responses of decision makers whose relevant characteristics may be known but whose responses are not. On the other hand, the probability framework of an OMLR is intuitively appealing.

Ordered Multinomial Logistic Regression Model

Following Greene's 1993 specification of the ordered choice models of Zavoina and McElvey, a decision maker's preference, y^*, could be viewed as a linear function of observable characteristics, x, of that decision maker, and some unobservable component, ε:

$$(1) \quad y^* = x\beta + \varepsilon.$$

We cannot observe y^* directly as it is a subjective measure of "degree of agreement" with the question asked. If we force the agent to choose only one of J possible ranked discrete choices, it seems reasonable to assume that there are thresholds of agreement, μ, segregating the agent's possible choices. In this case, if the discrete choices are numbered 1, 2, ..., J, we observe the following:

$$(2) \quad \begin{aligned} y &= 1 \quad \text{if } y^* \le 0, \\ &= 2 \quad \text{if } 0 < y^* \le \mu_1, \\ &= 3 \quad \text{if } \mu_1 < y^* \le \mu_2, \\ &\vdots \\ &= J \quad \text{if } \mu_{J-2} \le y^*. \end{aligned}$$

The μ's are parameters to be estimated along with β. Since the possible choices are ranked choices it must be the case that $0 < \mu_1 < \mu_2 < \ldots < \mu_{J-2}$. This model says that, given only J possible choices, an agent will choose the response that most closely represents his or her own feelings on the question.

We may normalize the mean and variance of ε in equation (1), to 0 and 1 respectively. If we also assume that the ε's originate from a symmetric probability distribution, then the probability version of equation (2) is

$$(3) \quad \begin{aligned} \text{prob}(y = 1) &= \Phi(-x\beta), \\ \text{prob}(y = 2) &= \Phi(\mu_1 - x\beta) - \Phi(-x\beta), \\ \text{prob}(y = 3) &= \Phi(\mu_2 - x\beta) - \Phi(\mu_1 - x\beta), \\ &\vdots \\ \text{prob}(y = J) &= 1 - \Phi(\mu_{J-2} - x\beta). \end{aligned}$$

The largest of the J probabilities, for a given agent's characteristics, x, is associated with the expected choice for that agent. The following marginal effects follow from differentiation of equation (3):

$$(4) \quad \begin{aligned} \frac{\partial \text{prob}(y=1)}{\partial x} &= -[\phi(-x\beta)]\beta, \\ \frac{\partial \text{prob}(y=2)}{\partial x} &= -[\phi(\mu_1 - x\beta) - \phi(-x\beta)]\beta, \\ \frac{\partial \text{prob}(y=3)}{\partial x} &= -[\phi(\mu_2 - x\beta) - \phi(\mu_1 - x\beta)]\beta, \\ &\vdots \\ \frac{\partial \text{prob}(y=J)}{\partial x} &= -[-\phi(\mu_{J-2} - x\beta)]\beta \end{aligned}$$

where ϕ is the first derivative of Φ. Again, for all of the probabilities to be positive, it must be the case that $0 < \mu_1 < \mu_2 < \ldots < \mu_{J-2}$.

One additional assumption regarding the probability distribution surrounding ε in equation (1) determines the specification of Φ. Two alternative specifications are often considered. If the disturbances are distributed standard normally, Φ denotes the cumulative distribution for the standard normal distribution, and ϕ denotes the probability density function for the standard normal distribution. In this case, the model specified as equations (3) and (4) is called a probit model. If the disturbances are distributed logistically, Φ is the logistic (or sigmoid) function, $\Phi(x) = 1/[1 + \exp(-x)]$, with the corresponding first derivative $\phi(x) = \Phi(x) \cdot [1 - \Phi(x)]$. In this case, the model becomes the OMLR model. Because the results of the two alternative models are typically quite similar (Amemiya), the lesser computational burden

estimating the logistic model often makes it preferred to the probit specification.[1]

Feedforward Backpropagation Neural Network Model

The neural network approach to choice modeling also considers that a decision maker's response to a particular question may be modeled as a function of observable characteristics of that agent, along with an unobservable component. The functional relationship depicted by a BPN model is substantially different from the OMLR-specified relationship. It is nonlinear, highly flexible, and has the form

$$(5) \quad y = f^L(\ldots f^2(f^1(f^0(\mathbf{x}\mathbf{W}_1)\mathbf{W}_2)\mathbf{W}_3)\ldots\mathbf{W}_{L+1}) + \varepsilon$$

where y denotes the discrete choice of the agent, $\mathbf{x}$ is a row vector of observed characteristics associated with the agent, and ε is some unobservable factor. The matrices $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \ldots, \mathbf{W}_{L+1}$ are matrices of parameters to be estimated. The functions $f^0, f^1, f^2, \ldots, f^L$ (superscripts notational, not exponential) are arbitrary monotonic nondecreasing transfer functions that map each matrix element in their respective arguments into a bounded space. Value L is the arbitrary number of hidden layers in the network model. Functional flexibility can be observed by the fact that the model in equation (5), even with only one hidden layer ($L = 1$), is a universal function approximator (Hornik, Stinchcombe, and White).

Essentially, the model depicted in equation (5) takes model predictions from one layer, transforms them through an f function, and then uses the transformed predictions as inputs to the next layer (a feedforward network). The dependent variable, y, could be thought of as a single output variable which takes on integer values of 1, 2, 3, ..., J, where J is the same as in the OMLR model—the number of alternative responses to a question modeled. Then, equation (5) represents a single equation model. Alternatively, a multiple-output equation system with J equations could be considered. In this case, the associated y's, $y_1, y_2, y_3, \ldots, y_J$, take on values of either 1 or 0, according to whether the subscript was the numbered response observed for the agent. The only change in equation (5) required to depict the latter approach is that y becomes a J-wide row vector, instead of a scalar, and $\mathbf{W}_{L+1}$ now has J columns rather than one column. In practice, a BPN model rarely involves more than one or two hidden layers and the f's associated with layers are often the same throughout. Also, in practice, the f chosen for the model is often the same logistic function of the OMLR, $f(x) = 1/(1 + e^{-x})$.

While the number of $\mathbf{W}_1$ rows and $\mathbf{W}_{L+1}$ columns are predetermined by the number of input and output variables respectively, the number of columns in each of the other $\mathbf{W}$ matrices is user-determined and referred to as the number of nodes in the corresponding hidden layer. Associated row dimensions follow from conformability. Like the OMLR model, the BPN model requires a constant term, called a bias in BPN parlance. Inclusion of a bias term typically occurs at each layer. In equation (5), to accommodate the bias terms, we should consider the $\mathbf{x}$-vector, and each of the transformed hidden layer prediction vectors, as having a 1 appended. Alternatively, we can make explicit the additional bias weights estimated. Doing just that, along with keeping f the same function throughout, for an n-observational data set, equation (5) is estimated as

$$(6) \quad \mathbf{Y} = f(\ldots f(f(f(\mathbf{X}\mathbf{W}_1 + \mathbf{i}_n\mathbf{w}_1^T)\mathbf{W}_2 + \mathbf{i}_n\mathbf{w}_2^T)\mathbf{W}_3 + \mathbf{i}_n\mathbf{w}_3^T)\ldots\mathbf{W}_{L+1} + \mathbf{i}_n\mathbf{w}_{L+1}^T) + \mathbf{E}.$$

If h_0 is the number of input variables, $\mathbf{X}$ is an $n \times h_0$ matrix of input data. Similarly, $\mathbf{Y}$ is an $n \times h_{L+1}$ matrix of output data, where h_{L+1} is the number of output variables, and L again denotes the number of hidden layers in the network model. Matrix $\mathbf{W}_1$ is an $h_0 \times h_1$ matrix of parameters (weights) to be estimated, with h_1 being the user-defined number of nodes in the first hidden layer. The notation $\mathbf{i}_n$ is an $n \times 1$ vector of 1's, and $\mathbf{w}_1$ is an $h_1 \times 1$ vector of weights to be estimated (accommodating the bias at that layer). Dimensions of successive

[1] The marginal approach depicted in equation (4), although routinely seen in empirical work, may not always be appropriate. First, Greene suggests that it is not appropriate for evaluating the effect of a dummy explanatory variable. Second, a small change in an explanatory variable may reinforce the predicted response more than an alternative choice, leaving the predicted response change indeterminate (since the predicted response is merely the largest of the predicted probabilities). Third, the model-predicted change-in-probability associated with a "more distant" choice (from the predicted response) may be smaller or larger in absolute value than the change-in-probability associated with a "closer" choice, implying a potential discontinuity in the predicted response associated with a small change in an explanatory variable. These problems are partly why we have used direct examination of predicted responses rather than partial derivatives to get at the underlying economics.

W's and **w**'s are similarly determined. Matrix transpose is denoted by T. Matrix **E** is an $n \times h_{L+1}$ matrix of disturbances.

Empirical Applications of Ordered Logit and Feedforward Backpropagation Models: Questions Analyzed and Explanatory Variables

A survey of 1,963 members of the Kansas Farm Management Association was administered in the fall of 1992. The 572 usable responses were matched to a detailed set of farm management records for each enterprise. The survey collected information on demographic characteristics of farmers, policy and crop insurance preferences, as well as risk attitudes. Five questions or statements solicited self-assessed risk preference information. A five-point Likert-type response scale for four questions allowed the following responses (mapped to integers 1 through 5, respectively): strongly agree, agree, unsure, disagree, and strongly disagree. The fifth question allowed responses ranging from 1 (risk hating) to 10 (risk loving), and were mapped to 1 through 5 to allow modeling with the same five-choice models applied to the first four questions. The responses to the questions are described in table 1, along with the frequency and proportion of each categorical response.

Twenty-eight variables were chosen from either survey information or the associated farm management information to serve as explanatory variables. All 28 variables were used in each model.[2] The data set associated with each of the five questions was handled in the following manner. First, to remove possible bias associated with the order of the data, a random sort was completed, and then held constant across models. Second, each data set was sorted roughly into thirds, with the first two-thirds allowing for model building (and in-sample testing if applicable), and the last third used only for out-of-sample (OOS) prediction. Thus, each question has unique in-sample and out-of-sample data sets.

[2] Because of missing data, different numbers of observations were available for modeling each of the five questions (ranging from 437 on question 5 to 454 on question 3). As a point of reference, table 2 displays the sample statistics for the 458 farms that had observations for each of the 28 variables and answered at least one of the five questions analyzed.

Measurement Statistics for Comparing Prediction Accuracy Across Models

Each of the questions studied has a possibility of five discrete choices. Although many prediction accuracy test statistics exist, we report the following six measures; the first four are explicitly loss-function related, while the last two are generally accepted forecast accuracy test statistics. CPR (correct predictions) is the proportion of total responses predicted exactly accurately. BOT (bottom) is the proportion of responses which were actually a 1 or a 2 that were predicted as either a 1 or a 2. Similarly, TOP is the proportion of responses which were actually a 4 or a 5 that were predicted as either a 4 or a 5. BOTH is the proportion of responses which were actually a 1, 2, 4, or 5 that were predicted as either a 1 or a 2 (if actually a 1 or 2), or predicted as either a 4 or a 5 (if actually a 4 or 5). The relevance of BOT is that, for some issues, predicting responses of individuals expected to agree (SA or A) is important. TOP is the disagree counterpart. BOTH might be used if a modeler is interested in segregating both those who are expected to agree and those who are expected to disagree from those who may only be uncertain. The same intuition can be applied to question 5 even though it does not deal with agreement/disagreement. MAE (mean absolute error) is the average absolute error across all predictions, where absolute error is abs(actual response-predicted response). Similarly, RMSE is the square root of the average squared error. Both MAE and RMSE test statistics consider the "closeness" of a prediction. For example, predicting an A response as an SA is better than predicting an A as a D.

Baseline Models

Suppose a response forecaster had no model to effectively use known explanatory characteristics of an individual to predict that individual's response, but that the forecaster was privy to actual response data for other individuals. Intuitively, the modeler might predict an unknown response as the response associated with a random draw from the set of known responses. Although naive, this approach would at least account for expected frequency of response, and thus could at least provide a baseline forecast.

The prediction approach just discussed was applied to make predictions for the last third of

Table 1. Questions Asked, Frequency, and Proportion of Each Response Over Total Data Set

Question #1: My farm is protected from risk by geographic or other diversity.					
SA	A	U	D	SD	Total
9 / 0.02	102 / 0.23	126 / 0.28	181 / 0.40	31 / 0.07	449 / 1.00
Question #2: My risks are better addressed by specific-peril crop insurance than by multiple peril crop insurance.					
SA	A	U	D	SD	Total
17 / 0.04	89 / 0.20	164 / 0.37	158 / 0.36	17 / 0.04	445 / 1.00
Question #3: My farm enterprise is self-insured.					
SA	A	U	D	SD	Total
17 / 0.04	69 / 0.15	75 / 0.17	232 / 0.51	61 / 0.13	454 / 1.00
Question #4: I enjoy taking business risks.					
SA	A	U	D	SD	Total
5 / 0.01	102 / 0.23	46 / 0.10	236 / 0.52	62 / 0.14	451 / 1.00
Question #5: Rate your preference for risk and uncertainty by circling a number between 1 and 10, with 1 being risk hating and 10 being risk loving.					
1&2[a]	3&4	5&6	7&8	9&10	Total
56 / 0.13	100 / 0.23	201 / 0.46	71 / 0.16	9 / 0.02	437 / 1.00

Notes: SA, A, U, D, SD correspond to strongly agree, agree, unsure, disagree, and strongly disagree. The first number in a cell corresponds to frequency, the latter to proportion (some rounding errors).

[a] The original ten possibilities were transformed to five, to facilitate comparisons with other questions.

data, the out-of-sample (OOS) data set, associated with each of the questions analyzed. For each prediction a random draw was made, sampling with replacement, from the known responses in the first two-thirds of the data. Five thousand such runs were implemented for each of the five questions examined. Comparison of actual responses in the OOS data with predicted responses allowed for calculation of each of the six test statistics after each run. After sequentially sorting the 5,000 resultant values for a test statistic, the endpoints of a 95% confidence interval would be ordered observations 126 and 4,875.[3] If an alternative predictive model generates a test statistic that is outside that test statistic's confidence interval, then that alternative predictive model is statistically superior or inferior (at least regarding that test statistic) to the corresponding NAIVE model.

Empirical OMLR Models

Maximum likelihood was used to fit the OMLR model specified in equation (3) to the in-sample data (first two-thirds) associated with question 1. The estimated model, along with the explanatory data of the OOS set (last one-third) associated with question 1, was then used to make response predictions for the OOS data set. The same two-step procedure was then applied to questions 2, 3, 4, and 5, in turn. It should be noted that there is no a priori basis for using any particular test statistic in evaluating OMLR models in out-of-sample prediction. Greene notes that even in sample, unlike ordinary least squares regression, an OMLR model does not seek to minimize the sum of squared prediction errors.

Empirical BPN Models

In theory, the feedforward functional form of equation (6) could be estimated with various least squares nonlinear optimization techniques,

[3] This method was used to preclude the possibility of a two-standard deviation lower confidence limit from falling below 0, which is implausible since four of the test statistics are bounded on the bottom by 0. In actuality, the two-standard deviation approach generated very similar results.

well-established or novel.[4] In practice, because of potential convergence and local minima problems, an algorithm unique to the neural networks field, called backpropagation, is used instead. Prior to a brief description of this algorithm, which gives the BPN its name, it may be helpful to display an estimated version of equation (6) in a recursive format to better visualize the "layers." The usual associated-with-prediction "hats" on the **W**'s, **w**'s, and **E** have been suppressed, but retained on the $\hat{\mathbf{Y}}$ to make it clear that these are predicted values. This recursive model (with the **P**'s as stage-wise predictions) is specified as

$$
\begin{aligned}
(7)\quad & \mathbf{P}_0 = \mathbf{X} \\
& \mathbf{P}_1 = f(\mathbf{P}_0\mathbf{W}_1 + \mathbf{i}_n\mathbf{w}_1^T) \\
& \mathbf{P}_2 = f(\mathbf{P}_1\mathbf{W}_2 + \mathbf{i}_n\mathbf{w}_2^T) \\
& \quad\vdots \\
& \mathbf{P}_{L+1} = f(\mathbf{P}_L\mathbf{W}_{L+1} + \mathbf{i}_n\mathbf{w}_{L+1}^T) \\
& \hat{\mathbf{Y}} = \mathbf{P}_{L+1} \\
& \mathbf{E} = \mathbf{Y} - \hat{\mathbf{Y}}.
\end{aligned}
$$

As in the OMLR model, the model-predicted output for a particular agent is not ordinarily an integer from 1 to J as it needs to be to effect a predicted response for that agent. But also like in the OMLR model, if **Y** is modeled as a J-output system, the largest of the predicted output variable values, for a given agent, can be associated with the predicted response for that agent. On the other hand, if **Y** is modeled as a single output variable, the integer (1, 2, ..., J) that is closest numerically to the predicted output value is considered the model-predicted response.

The BPN estimation algorithm uses an iterative gradient descent approach in seeking to minimize the model sum of squared errors in equation (7) (or the sum of the diagonal of $\mathbf{E}'\mathbf{E}$ if **Y** is modeled as J outputs). The iterative weight-adjustment process begins by scaling the output error by the first derivative of the f transfer function, evaluated at the output layer's inputs (previous layer's transformed predictions or output), to become the output layer's error information term, $\mathbf{D}_{L+1}$:

$$(8)\quad \mathbf{D}_{L+1} = \mathbf{f}'\#^* (\mathbf{Y} - \mathbf{P}_{L+1})$$

where $\mathbf{f}'$ is the first derivative of f, and #* denotes element-by-element matrix multiply. If f is the logistic (sigmoid) specification, equation (8) becomes

$$(9)\quad \mathbf{D}_{L+1} = \mathbf{P}_{L+1}\#^* (\mathbf{i}_n\mathbf{i}_{h_{L+1}}^T - \mathbf{P}_{L+1})\#^* (\mathbf{Y} - \mathbf{P}_{L+1})$$

where **i** denotes a column vector of ones whose length is depicted by its subscript. The gradient descent procedure implies that $\mathbf{W}_{L+1}$ and $\mathbf{w}_{L+1}$ could immediately be updated by adding to their elements terms derived by scaling $\mathbf{D}_{L+1}$ by the inputs to the $L + 1$ layer. But, the nonupdated $\mathbf{W}_{L+1}$ and $\mathbf{w}_{L+1}$ are first needed to compute the updated terms for the weight matrices of the previous layers.[5]

Given the error information matrix for the output layer, $\mathbf{D}_{L+1}$, and the current output layer weight values, $\mathbf{W}_{L+1}$, an expected error matrix for the immediately preceding layer is implied (since we do not know the true error as we do in the output layer), and so on. Continuing with the example:

$$
\begin{aligned}
(10)\quad & \mathbf{D}_L = \mathbf{P}_L\#^* (\mathbf{i}_n\mathbf{i}_{h_L}^T\mathbf{P}_L)\#^* [\mathbf{D}_{L+1}\mathbf{W}_{L+1}^T] \\
& \mathbf{D}_1 = \mathbf{P}_1\#^* (\mathbf{i}_n\mathbf{i}_{h_1}^T - \mathbf{P}_1)\#^* [\mathbf{D}_2\mathbf{W}_2^T]
\end{aligned}
$$

where expected errors for each layer are in the brackets.[6] The **i**-subscripts imply h_1 nodes in the first hidden layer, and h_L nodes in the last hidden layer. The weight matrices are then updated as

$$
\begin{aligned}
(11)\quad & \mathbf{w}_{L+1} = \mathbf{w}_{L+1} + \mathbf{D}_{L+1}^T\mathbf{i}_n, \\
& \mathbf{W}_{L+1} = \mathbf{W}_{L+1} + \mathbf{P}_L^T\mathbf{D}_{L+1}, \\
& \mathbf{w}_L = \mathbf{w}_L + \mathbf{D}_L^T\mathbf{i}_n, \\
& \mathbf{W}_L = \mathbf{W}_L + \mathbf{P}_{L-1}^T\mathbf{D}_L, \\
& \quad\vdots \\
& \mathbf{w}_1 = \mathbf{w}_1 + \mathbf{D}_1^T\mathbf{i}_n, \\
& \mathbf{W}_1 = \mathbf{W}_1 + \mathbf{P}_0^T\mathbf{D}_1.
\end{aligned}
$$

[4] A particularly promising algorithm may be the Levenberg-Marquardt algorithm (More; Hagan and Menhaj). For especially difficult-to-solve estimations, genetic algorithms may be promising as well (Dorsey and Mayer).

[5] Thus, the proper sequence must be observed. Notice that the $L + 1$ subscript on P implies that the derivative is evaluated at that layer's transformed output rather than that layer's input, as it should be. But recall that the first derivative of the logistic function evaluates to a function which has as its argument the logistic function itself.

[6] Alternatively, if the expected layer error framework is not appealing, equations (9) and (10) follow directly from differentiation of a squared output error function, $1/2\ E^2$ [E, as in equation (7)], with respect to each weight to be estimated (the W's and w's). For a rigorous mathematical derivation of the backpropagation estimation algorithm, the reader should consult Kosko.

Although BPNs are feedforward networks in that layer predictions move forward through the network, equations (9) and (10) show that the error information terms start at the output layer and propagate backwards through the preceding layers—hence backpropagation. In practice, full updates are not typically made. That is, the second terms on the right-hand side of equation (11) are usually multiplied by an adjustment factor (the learning coefficient), α, where $0 < \alpha < 1$. Also, a portion of the update made in the previous iteration is often added to the current update, in an effort to prevent getting trapped in a local minimum. This portion, η, is called a momentum factor.

Many variants on equations (9) through (11) may be considered as well. The method depicted above is called batch training, where network weights are adjusted after all data have been presented. In this sense, ordinary least squares regression is batch trained. An alternative to batch training is on-line training, where the network weights are updated after each pattern presentation, and where a pattern is the BPN term for one observation on the data—a single row vector in **X**, along with the associated row vector in **Y**. Additional variants may relate to α and η. For example, each layer may have its own α or η. Since smaller weight adjustments tend to occur at preceding layers in a BPN, larger learning coefficients can be used with the smaller adjustments to accelerate the estimation process while still not causing undue network weight oscillation. Both α and η may be gradually decreased after a preselected number of iterations as well.

Besides the functional form and estimation algorithm decisions required of the modeler, other decisions relate to initialization and stopping points of the iterative weight adjustment process, as well as scaling of the data. For a brief discussion of some of these possibilities see Kastens, Featherstone, and Biere, or any textbook on neural networks (e.g., Fausett).

Because BPN choice models are not supported by a large body of empirical research, there is little a priori basis for choosing a particular BPN (given the explanatory variables) for effecting out-of-sample prediction. We follow the BPN convention of estimating alternative models based on a portion of the in-sample data (in our case the first one-half, which is the first one-third of the total data associated with a question), and testing the alternative models over the balance of the in-sample data. The model that performs the best on the in-sample testing data (the middle one-third of the total data) is then used to predict the final OOS responses. Each of the six prediction accuracy test statistics discussed earlier is used to define best, in turn.

The infinite number of models that could have been examined was narrowed to a manageable number as follows. The algorithmic solution process for a handful of arbitrary modeling parameters was observed over several iterations (only question 5 data was used here) to visually check for a tendency for in-sample mean squared model errors to more-or-less monotonically decrease over increasing iterations. Then, if pertinent, a small discrete range for each such examined modeling parameter was included in the set of total modeling parameters considered, in determining the final BPNs used to make the out-of-sample predictions. We settled on estimating 192 unique alternative BPNs using the in-sample training data (first third of total data). The best of the 192 models in predicting in-sample testing set (middle third) responses, using one of the test statistics, was then used to predict the responses in the OOS (last third) data.

Responses were modeled two ways: first, as a single output variable that can take on values of 1, 2, 3, 4, or 5; and second, as five output variables where, for a single observation on the data, the output variable whose subscript matches the response chosen is valued at 1, and all other output variables for that observation are valued at 0. Two transfer functions were considered: (*a*) the logistic or sigmoid, $f(x) = 1/(1 + e^{-x})$, with the corresponding first derivative, $f'(x) = f(x) * [1 - f(x)]$; and (*b*) the hyperbolic tangent, $f(x) = (e^{x} - e^{-x})/(e^{x} + e^{-x})$, with the corresponding first derivative, $f'(x) = 1 - [f(x)]^2$.

As is conventional for the two transfer functions, if the logistic function was chosen, input data were scaled between 0 and 1 prior to use in the network; if the hyperbolic tangent function was chosen, the input data were scaled between –1 and 1. For example, in the latter case, the maximum value over the in-sample data, for a particular input variable, was mapped to 1 and the minimum to –1. All other observations on that variable were linearly mapped to values between –1 and 1. It would not be appropriate to use maximum and minimum information from the OOS data if we think of that demographic farm/farmer data as not being available until chronologically after the predictive models are constructed. Thus, OOS input variables

Table 2. Explanatory Variables (Farm/Farmer Characteristics) Used in the Prediction Models

Variable Description (through, or as of end of 1991)	Mean	Std. Dev.
Total farm acres (all operated, including rented and pasture)	1,577.7860	1,280.9882
Proportion of farm that is devoted to crop production	0.7186	0.2485
Proportion of total acres that are rented	0.6250	0.3012
Debt/assets ratio for farm	0.4161	0.4171
Farm net worth ($1,000)	362.5833	374.3466
Farm net worth per acre ($/acre)	332.4195	440.1186
Rate of return to farm net worth (%)	5.7064	371.9417
Change in net worth beginning to end of 1991 ($1,000)	7.9907	51.3181
Change in net worth per acre ($/acre)	9.7326	69.2942
Net farm income ($1,000)	25.7953	50.4635
Government payments ($1,000)	18.8951	15.2998
Government payments per acre ($/acre)	13.3635	8.5091
Gross value of crop production, including landlord share ($1,000)	127.6633	122.7317
Gross value of crop production per acre ($/acre)	123.8606	53.6314
Crop production costs on accrual basis ($1,000)	77.1037	64.2486
Crop production costs per acre ($/acre)	85.0931	43.3537
Cropping efficiency (gross value/costs in crop production)	1.7582	1.5898
1 if farm is a corporation, 0 otherwise	0.0349	0.1838
1 if farm is a partnership, 0 otherwise	0.0655	0.2477
1 if farm is primarily a crop farm, 0 otherwise	0.6201	0.4859
1 if farm is primarily a livestock farm, 0 otherwise	0.0808	0.2728
Total number of family dependents	3.1659	1.5411
Age of primary operator	50.0524	12.4957
Years of formal education of operator (12=high school etc.)	13.9443	2.0919
Miles farm is located from nearest town	7.5835	13.7927
1 if recently have been a marketing club member, 0 otherwise	0.1550	0.3623
1 if believed price risk more important than yield risk, 0 otherwise	0.6092	0.4885
1 if purchased multiperil crop insurance in 1991, 0 otherwise	0.6310	0.4831

were scaled using the same maximums and minimums from the in-sample data, and so may take on values greater than 1 or less than –1. As suggested by NeuralWare, a commercial neural network software provider, output data were scaled between –0.85 and 0.85 (when using the hyperbolic tangent function) or between 0.15 and 0.85 (when using the logistic function). All actual and predicted output values were transformed to their original scales before calculation of the various test statistics.[7]

Starting values of network weights to be estimated were uniformly random values between –1 and 1. At initialization, to prevent starting bias in the weights, the randomly selected weights were normalized to unit-vector length, dividing the elements in each weight matrix column (here the associated bias weight was also included as a "column" member) by the Euclidean norm for that column. Batch training was used throughout. An iteration consisted of one batch presentation (all of the training data) along with the subsequent network weight updates. Two different stopping points were considered for the learning process, after 1,000 iterations and after 3,000 iterations.

Modeling behavioral data should include only one or two hidden layers and only a few nodes per hidden layer (Fausett). Models were considered with either one or two hidden layers, with either 3 or 5 nodes in the first hidden layer. When the second hidden layer was included, it contained one less node than the first hidden layer. Two possibilities for the momentum factor, η, were considered: 0.4 and 0.6. Three possibilities for the learning coefficient, α, used at the first hidden layer were considered, depending upon the choice of transfer function. When the logistic transfer function was applied, the three alternatives for α were

[7] The arbitrary scaling over [–0.85, 0.85] or [0.15, 0.85] is to allow the model to make output predictions outside the range of the in-sample output data (since the logistic and hyperbolic transfer functions are bounding functions). Alternatively, this could be accomplished by using a linear transfer function in the last layer. In that case, another hidden layer may have to be added to capture the same range of underlying nonlinearities.

0.1, 0.2, and 0.3. With the hyperbolic tangent function, the three α possibilities were 0.001, 0.002, and 0.003. The α for each successive layer was always established as 0.83 times the α for the preceding layer. The model parameters (α, η) were held constant during the estimations. While the coding for the backpropagation solution process could have been written in virtually any numerical or statistical software, we used Matlab Numeric Computation Software because of its convenience as a matrix language program.

Results and Discussion

The expected test statistics and associated confidence intervals for each of the five baseline (NAIVE) models are reported in table 3. The table shows the expected test statistics based upon the process of randomly drawing from the known responses associated with the in-sample data. For example, if we predict OOS responses to question 1, based upon this naive process, we would expect to predict 29.73% exactly correctly (CPR). We would expect to correctly determine those individuals who feel strongly (predicting those who respond 1 or 2 as either 1 or 2, and those who respond 4 or 5 as either 4 or 5) 39.85% of the time (BOTH).

To illustrate the test statistics, the prediction matrix and test statistics for the OMLR model associated with question 1 are reported in table 4. The OMLR model predicted all three of the 1 responses as a 4 response. Of the 30 "2" responses, 7 were correctly predicted as a 2 and 23 were predicted as a 4. The correct predic-

Table 3. Expected Test Statistics and Associated Confidence Intervals for NAIVE Models in Out-of-Sample Testing Data Set Predictions

	Out-of-Sample Prediction Test Statistics for NAIVE Models					
	CPR	BOT	TOP	BOTH	MAE	RMSE
	Question #1 (148 responses)—Farm Protected from Risk by Geography					
Upper confidence	0.3716	0.4242	0.5753	0.4906	1.1689	1.4727
Expected value	0.2973	0.2585	0.4618	0.3985	1.0397	1.3506
Lower confidence	0.2297	0.1212	0.3562	0.3113	0.9122	1.2247
	Question #2 (147 responses)—Risks Better Addressed with Specific Peril Ins.					
Upper confidence	0.3673	0.4062	0.4706	0.4200	1.1293	1.4333
Expected value	0.2986	0.2485	0.3581	0.3230	1.0075	1.3145
Lower confidence	0.2313	0.0938	0.2500	0.2300	0.8912	1.1924
	Question #3 (150 responses)—Farm Is Self-Insured					
Upper confidence	0.4000	0.4091	0.7228	0.6423	1.2000	1.5684
Expected value	0.3318	0.2124	0.6312	0.5563	1.0573	1.4290
Lower confidence	0.2667	0.0455	0.5347	0.4715	0.9200	1.2832
	Question #4 (149 responses)—Enjoy Taking Business Risks					
Upper confidence	0.4362	0.3939	0.7333	0.6232	1.2013	1.5630
Expected value	0.3635	0.2463	0.6401	0.5459	1.0524	1.4263
Lower confidence	0.2886	0.1212	0.5524	0.4710	0.9128	1.2875
	Question #5 (144 responses)—Self-Ranked Risk Preference					
Upper confidence	0.3750	0.4576	0.3913	0.3902	1.1528	1.4790
Expected value	0.3025	0.3324	0.1961	0.2942	1.0261	1.3464
Lower confidence	0.2361	0.2203	0.0435	0.1951	0.9028	1.2162

Note: Expected values based on 5,000 random samples (sampling with replacement) from the responses associated with the in-sample data set. Confidence intervals (95%) were established using endpoints associated with ordered responses numbered 126 and 4,875.

Table 4. Prediction Matrix of the OMLR Model Used for Question 1

		Predicted No. of Times Response Below Was Chosen					Row Sums
		1	2	3	4	5	
Actual No.	1	0	0	0	3	0	3
of Times	2	0	7	0	23	0	30
Response	3	0	4	0	38	0	42
at Right	4	0	7	0	54	0	61
was Chosen	5	0	1	0	10	1	12
Column Sums	>	0	19	0	128	1	148

Note: CPR = 0.4189, BOT = 0.2121, TOP = 0.8904, BOTH = 0.6792, MAE = 0.8378, RMSE = 1.1855.

tions are on the diagonal; so CPR is the diagonal sum divided by the total responses (62/148). BOT is the sum of the upper left-hand four elements divided by the sum of the upper two elements in the Row Sums column (7/33). TOP is the sum of the lower right-hand four elements divided by the sum of the lower two elements in the Row Sums column (65/73). BOTH is the sum of the numerators in BOT and TOP, divided by the sum of their denominators (72/106). Five out of six of the test statistics (all but BOT) suggest that the OMLR model is statistically superior to the naive baseline model for question 1. That is, they are outside the confidence interval, and on the preferred side (we want CPR, BOT, TOP, and BOTH to be above the expected values in the naive model, and MAE and RMSE to be below).

At first glance, it appears that the OMLR model may take its strength from overpredicting those responses that make up a large portion of the possible responses. That is, many responses were a 4 (61 out of 148); but the model extended this by predicting 128 out of the 148 as a 4. If a hypothetical loss function is such that large penalties are attached to "missing" those that respond with a 4 (that is, if the penalty associated with type I error is low), then the OMLR appears to have done well. But, what if a modeler's interest is in determining correctly those who agree with question 1 (the 33 responses that were either a 1 or a 2)? Here, the naive model is expected to correctly determine 25.85% of those individuals while the OMLR model only correctly determined 21.21%. Nominally, but not statistically, it appears that in this case the modeler would be better off ignoring the information on the 28 characteristics associated with a respondent and just use the NAIVE approach. This problem may be especially keen for an OMLR if many responses are "unsure" and if a modeler is really not interested in the unsure responses.[8]

For each question examined we considered six alternatively chosen BPN models. The first BPN used for predicting out-of-sample responses was the one (of the total 192 possible) which had performed the best, using the CPR criterion, in predicting the responses associated with the in-sample testing set (the middle third of the total data). Similarly, each of the succeeding BPN models followed from using one of the other criteria in turn. The functional form for each of the BPN models selected using these procedures is described in table 5. In addition, the control parameters for the backpropagation algorithm used to fit each functional form are also listed in table 5.

Most (21 of 30) of the selected models depict the response variable in a five-output system rather than in a single-output equation. This is similar to the econometric suggestion that the logit model should be superior for a J-classification problem than would be ordinary least squares regression using one dependent variable which takes on values 1, 2, ..., J. Only 9 of 30 models used the hyperbolic tangent transfer function. While no explanation for this is immediately apparent, it does lend credence to

[8] A reviewer suggested that the prediction accuracy of the OMLR may have been hindered by multicollinearity, and, hence, that a more parsimonious model (using less explanatory variables) may be superior. The 28 explanatory variables were included in each model (logistic regression or neural network) because each was thought to contribute something unique to explaining risk response. Among the 378 unique pairwise linear correlations within the explanatory data, only 8 were in the range of 0.50. Nonetheless, to examine whether a more parsimonious OMLR model may be a better predictor, we estimated OMLRs using only the 14 variables (of the 28) thought to be the most important explanatory variables. The associated out-of-sample predictions were generally less accurate, by any of the test statistics considered, than were predictions arising from the 28-variable models.

Table 5. BPN Model Structure and Estimation Parameters

Model	Functional Form				Estimation Parameters		
	No. of Output Nodes	Transfer[a] Function	No. of Hidden Layers	No. of Nodes in First Hidden Layer[b]	Momentum η	Learning α[c]	No. of Iterations
	Question #1—Farm Protected from Risk by Geography						
BPN-CPR	5	tanh	2	5	0.6	0.002	1,000
BPN-BOT	1	tanh	2	3	0.6	0.002	3,000
BPN-TOP	5	sig	1	3	0.4	0.3	3,000
BPN-BOTH	5	sig	1	3	0.4	0.3	3,000
BPN-MAE	1	sig	1	3	0.6	0.3	1,000
BPN-RMSE	1	sig	1	3	0.6	0.3	1,000
	Question #2—Risks Better Addressed with Specific-Peril Insurance						
BPN-CPR	5	sig	1	5	0.6	0.3	1,000
BPN-BOT	1	sig	1	3	0.6	0.3	1,000
BPN-TOP	5	sig	2	5	0.4	0.3	1,000
BPN-BOTH	5	sig	2	5	0.6	0.3	3,000
BPN-MAE	5	sig	2	3	0.6	0.3	3,000
BPN-RMSE	5	sig	1	3	0.6	0.3	1,000
	Question #3—Farm Is Self-Insured						
BPN-CPR	5	tanh	2	3	0.6	0.003	1,000
BPN-BOT	5	sig	1	3	0.6	0.3	1,000
BPN-TOP	5	sig	1	3	0.4	0.1	1,000
BPN-BOTH	5	tanh	2	3	0.6	0.003	1,000
BPN-MAE	5	tanh	1	3	0.6	0.003	1,000
BPN-RMSE	5	sig	1	3	0.6	0.2	1,000
	Question #4—Enjoy Taking Business Risks						
BPN-CPR	5	sig	2	5	0.6	0.3	1,000
BPN-BOT	5	tanh	1	5	0.4	0.002	3,000
BPN-TOP	5	sig	1	3	0.6	0.3	1,000
BPN-BOTH	5	sig	2	5	0.6	0.3	1,000
BPN-MAE	1	tanh	2	3	0.4	0.001	1,000
BPN-RMSE	1	sig	2	5	0.4	0.1	1,000
	Question #5—Self-Ranked Risk Preference						
BPN-CPR	5	sig	1	5	0.6	0.2	3,000
BPN-BOT	1	tanh	1	3	0.6	0.003	1,000
BPN-TOP	1	sig	1	5	0.6	0.3	1,000
BPN-BOTH	1	tanh	1	3	0.6	0.003	1,000
BPN-MAE	5	sig	2	3	0.4	0.3	3,000
BPN-RMSE	5	sig	2	3	0.4	0.3	3,000

[a] The sigmoid (logistic) function is denoted by sig and the hyperbolic tangent function by tanh.
[b] When a second hidden layer is present it contains one less node than the first hidden layer.
[c] The learning coefficient is for the first hidden layer. The learning coefficient for each succeeding layer is 0.83 times that of the preceding layer.

the common use of the sigmoid transfer function in the neural networks literature.

Generally, the selection process chose models that were simpler in design then they might have been. That is, 17 of the 30 models used only one hidden layer and 20 of the 30 models used only three nodes in the first hidden layer. This illustrates the well-known econometric principal that overfitting data to highly nonlinear functional forms may actually lead to worse predictions out-of-sample. In addition, 21 of the 30 models used only 1,000 iterations rather than 3,000. This illustrates the well-known neural network principal that overtraining in model estimation may also lead to worse predictions.

For each question, the test statistics for the six BPNs, along with the test statistics associated with the corresponding NAIVE (repeated from table 3 to ease comparison) and OMLR models are reported in table 6.[9] Several items are noteworthy regarding the results reported in table 6. First, responses for some questions may be easier to predict than for others. For question 3 ("My farm is self-insured") only 3 of the 42 test statistics display models that are not statistically superior to the corresponding NAIVE models. Additionally, none of the test statistics indicate models statistically worse than corresponding NAIVE models. For this question, given the choice of a forecast accuracy test statistic, the models generally appear similar in predictive accuracy. OMLR, BPN-CPR, BPN-BOT, and BPN-BOTH all performed exceptionally well for this question, regardless of the choice of test statistic. Perhaps farmers' perceptions regarding self-insurance are uniquely determined by the levels of the various explanatory characteristics pertaining to those farmers. On the other hand, the tabular components associated with each of the other four questions typically display several models which were actually worse statistically than the associated NAIVE models.

Table 6 suggests that the BOT statistic is associated with greater predictive difficulty relative to the other statistics. Many models in the BOT column appear worse than corresponding NAIVE models. Twelve out of 35 are even statistically inferior to the NAIVE model. Recall that BOT deals with predicting accurately the responses of those who agree or strongly agree with the questions. Table 1 shows relatively few responses in the strongly agree categories, and even the sum of SA and A tends to be much less than the sum of D and SD. It may be that the data are simply not rich enough for the models to adequately distinguish response categories that have fewer responses.

One of our objectives was to compare the standard econometric model (OMLR) with neural network models in out-of-sample prediction accuracy. Ignoring NAIVE, table 6 displays 171 pairwise comparisons involving OMLR (9 of the 180 possible are indeterminate due to equal test statistic values). In 114 of the 171 comparisons, the OMLR test statistic was better than that of the competing model (no statistical judgment made in better, only that the test statistic was larger—in the case of CPR, BOT, TOP, or BOTH—or smaller—in the case of MAE or RMSE). This implies that the OMLR approach was superior to the neural network competition 0.6667 of the time, although we might expect it to be better only 0.5 of the time. Based upon a one-tailed t-test, 0.6667 is significantly larger than 0.5 at the 99.99% confidence level. Similar analysis involving each of the BPN models in turn (compared pairwise with OMLR and the other BPN models) showed that none of the associated proportions was significantly larger than 0.5 at the 95% confidence level.

The foregoing provides evidence in support of using the relatively less computationally intensive OMLR model over the BPN models for out-of-sample survey response prediction. But, this statement is highly conditional upon indifference to the choice of which test statistic is to be used in forecast accuracy comparison. For many problems, the choice of test statistic is not irrelevant. Two issues arise: (*a*) Is the test statistic compatible with the loss function of the modeler? and (*b*) Is the test statistic consistent? That is, when the test statistic is used to rank alternative models over some in-sample or out-of-sample data set, does it produce similar rankings on still further out-of-sample data? Regarding the first issue, loss function compatibility, as noted earlier, it may be that a modeler wishes to better identify farmers who feel more strongly (the BOTH criterion) than farmers in general. Along this line, when BOTH was the test statistic used, BPN-BOTH was superior to OMLR in three of five cases; and when BPN-BOTH was inferior to OMLR (questions 3 and 4) it still was a good model, statistically better than the corresponding NAIVE model. On the other hand, when OMLR was worse than BPN-BOTH (by the BOTH criterion), it tended to be substantially worse. In fact, in question 5 (self-

[9] The predictive accuracy test statistics are available from the authors upon request.

Table 6. Prediction Test Statistics for the NAIVE, OMLR, and BPN Models

	Out-of-Sample Prediction Test Statistics					
	CPR	BOT	TOP	BOTH	MAE	RMSE
	Question #1—Farm Protected from Risk by Geography					
NAIVE	0.2973	0.2585	0.4618	0.3985	1.0397	1.3506
OMLR	0.4189[b]	0.2121	0.8904[b]	0.6792[b]	0.8378[b]	1.1855[b]
BPN-CPR	0.2973	0.2727	0.4110	0.3679	0.9595	1.2302
BPN-BOT	0.2297[w]	0.3333	0.2877[w]	0.3019[w]	1.1419	1.4403
BPN-TOP	0.4122[b]	0.0000[w]	1.0000[b]	0.6887[b]	0.8311[b]	1.1654[b]
BPN-BOTH	0.4122[b]	0.0000[w]	1.0000[b]	0.6887[b]	0.8311[b]	1.1654[b]
BPN-MAE	0.2838	0.0000[w]	0.0000[w]	0.0000[w]	0.8176[b]	1.0101[b]
BPN-RMSE	0.2838	0.0000[w]	0.0000[w]	0.0000[w]	0.8176[b]	1.0101[b]
	Question #2—Risks Better Addressed with Specific-Peril Insurance					
NAIVE	0.2986	0.2485	0.3581	0.3230	1.0075	1.3145
OMLR	0.3810[b]	0.1250	0.3676	0.2900	0.7619[b]	1.0302[b]
BPN-CPR	0.3673[b]	0.1250	0.5147[b]	0.3900	0.8571[b]	1.1780[b]
BPN-BOT	0.2109[w]	0.6250[b]	0.0294[w]	0.2200[w]	1.0748	1.2936
BPN-TOP	0.4490[b]	0.0000[w]	0.8529[b]	0.5800[b]	0.7755[b]	1.1547[b]
BPN-BOTH	0.3197	0.1250	0.5294[b]	0.4000	0.9864	1.3015
BPN-MAE	0.4286[b]	0.0000[w]	0.5147[b]	0.3500	0.7279[b]	1.0400[b]
BPN-RMSE	0.3197	0.0000[w]	0.0000[w]	0.0000[w]	0.7823[b]	0.9932[b]
	Question #3—Farm Is Self-Insured					
NAIVE	0.3318	0.2124	0.6312	0.5563	1.0573	1.4290
OMLR	0.5600[b]	0.5455[b]	0.9703[b]	0.8943[b]	0.5467[b]	0.8869[b]
BPN-CPR	0.5400[b]	0.5455[b]	0.9208[b]	0.8537[b]	0.6000[b]	0.9592[b]
BPN-BOT	0.5133[b]	0.7273[b]	0.8218[b]	0.8049[b]	0.6800[b]	1.0646[b]
BPN-TOP	0.5133[b]	0.2727	0.9703[b]	0.8455[b]	0.6067[b]	0.9345[b]
BPN-BOTH	0.5400[b]	0.5455[b]	0.9208[b]	0.8537[b]	0.6000[b]	0.9592[b]
BPN-MAE	0.4933[b]	0.1818	0.9010[b]	0.7724[b]	0.6400[b]	0.9661[b]
BPN-RMSE	0.5533[b]	0.1818	0.9703[b]	0.8293[b]	0.5400[b]	0.8679[b]
	Question #4—Enjoy Taking Business Risks					
NAIVE	0.3635	0.2463	0.6401	0.5459	1.0524	1.4263
OMLR	0.5839[b]	0.1818	0.9619[b]	0.7754[b]	0.6242[b]	1.0199[b]
BPN-CPR	0.4832[b]	0.1818	0.8286[b]	0.6739[b]	0.8523[b]	1.2612[b]
BPN-BOT	0.3893	0.4242[b]	0.5810	0.5435	1.0805	1.4632
BPN-TOP	0.5705[b]	0.0000[w]	1.0000[b]	0.7609[b]	0.6510[b]	1.0459[b]
BPN-BOTH	0.4832[b]	0.1818	0.8286[b]	0.6739[b]	0.8523[b]	1.2612[b]
BPN-MAE	0.4832[b]	0.0606[w]	0.8857[b]	0.6884[b]	0.7047[b]	1.0459[b]
BPN-RMSE	0.5034[b]	0.0303[w]	0.8667[b]	0.6667[b]	0.6711[b]	1.0100[b]
	Question #5—Self-Ranked Risk Preference					
NAIVE	0.3025	0.3324	0.1961	0.2942	1.0261	1.3464
OMLR	0.3819[b]	0.1525[w]	0.0435	0.1220[w]	0.8333[b]	1.1547[b]
BPN-CPR	0.3681	0.3898	0.1739	0.3293	0.8403[b]	1.1456[b]
BPN-BOT	0.3333	0.6610[b]	0.0870	0.5000[b]	0.8611[b]	1.1426[b]
BPN-TOP	0.2778	0.1525[w]	0.2174	0.1707[w]	0.9792	1.2555
BPN-BOTH	0.3333	0.6610[b]	0.0870	0.5000[b]	0.8611[b]	1.1426[b]
BPN-MAE	0.3542	0.5085[b]	0.0870	0.3902[b]	0.8264[b]	1.1087[b]
BPN-RMSE	0.3542	0.5085[b]	0.0870	0.3902[b]	0.8264[b]	1.1087[b]

Note: Statistical superiority to the corresponding baseline model, at the 95% confidence level, is indicated with a superscript *b*, for *better*; statistical inferiority is indicated with a *w*, for *worse*. Superior prediction is associated with larger CPR, BOT, TOP, and BOTH test statistics, but smaller MAE and RMSE test statistics.

ranked risk preference), OMLR was bad enough (by the BOTH criterion) that it was statistically worse than the NAIVE model.

Consider an executive whose charge it is to use known farmer characteristics data to assist in targeting those farmers who are risk averse (so that they can be solicited with a risk-averse marketing plan perhaps), as well as those farmers who may be more risk loving (so that they can be targeted with a risky investment proposal perhaps). It may also be desirous that solicitation expenditures are not wasted on farmers who are more or less indifferent to risk. The OMLR model would be ill-advised in this case. It appears that the executive would be better off discarding all relevant farmer characteristic data and simply using the NAIVE model to make the necessary predictions. On the other hand, using the BPN-BOTH procedure would have allowed the executive to make 70% more correct predictions (by the BOTH criterion) than using the NAIVE model. As noted earlier, it may be that the OMLR model tends to overpredict those responses that are heavily represented in the sample. The question 5 response information in table 1 shows that many (46%) of the respondents were more or less indifferent to risk. Could this be the reason that the OMLR model delivers the best CPR statistic and the worst BOT, TOP, and BOTH statistics for question 5 in table 6? Does it primarily do a good job in predicting those farmers who are indifferent to risk? If so, then the OMLR may not be compatible with loss functions that emphasize other predictions.

The second issue regarding choice of test statistics is consistency, and is observed by examining the tabular column headed by the same test statistic as the one used in selection of a BPN model. For example, to examine consistency of the BPN-CPR model we look at the CPR column for each question, observing whether the BPN-CPR model tended to be a better predictor than competing BPN models. Generally, the BPN models display consistency. When a model was chosen based upon a specific in-sample testing set criterion, it tended to be at least as good (on the basis of the specified criterion) as alternatively selected BPNs in out-of-sample prediction. For two of the criteria (TOP and RMSE) the BPN selected based upon the criterion in-sample was consistently in the top-ranked position, based upon the same criterion when compared with alternatively chosen BPNs in out-of-sample prediction. On the average, across the five questions and across the six BPN models, a specifically chosen BPN was at least as good as 4.467 of the possible 5 competing BPN models. Additionally, none of the entries on the diagonal from BPN-CPR to BPN-RMSE for each question depicts models that are statistically (or even nominally) worse predictors than the corresponding NAIVE models.[10]

Consistency, as defined above, is not so clearly demonstrated when model estimation and testing take place over the same in-sample data. Considering the possibility that it may not be prudent to waste the middle third of the data for in-sample testing, we examined models constructed by using the first two-thirds of the data for both model estimation and model selection (results not shown). Here, for just one example, the BPN model chosen based on its low RMSE consistently produced an out-of-sample model that was actually higher in RMSE than the alternatively chosen BPN models. This may be an example of the well-known tradeoff in economic model-building between in-sample and out-of-sample prediction error. The importance of rigorous out-of-sample testing when using highly flexible nonlinear functional forms is clear.

The foregoing consistency of the model-selection test statistics is encouraging for the neural network approach to model building:[11] (*a*) Select a model prediction test statistic based upon the relevant loss function. (*b*) Split the data into three sections. (*c*) Estimate alternatively configured BPNs using the first data section. (*d*) Compute the prediction test statistic for each model estimated in step (*c*) based upon its predictive accuracy over the second data section. (*e*) Perform a grid search to select the optimal model based upon the chosen test statistic. (*f*) Use the optimally chosen model over the third data section, reporting prediction results based upon the selected test statistic or others.

[10] Robustness of the procedures developed and discussed in this analysis was tested by completing each analytical procedure with in-sample training and in-sample testing sets reversed. That is, models were estimated using the middle third of data, selected by predicting responses associated with the first third of the data, and ultimately used to predict the out-of-sample responses associated with the final third of the data. Many of the quantitative results and almost all of the qualitative results reported in this section did not change after using such a reversal process.

[11] This is not to say that the technique of adaptively selecting alternative forecasting models is unique to neural networks (e.g., Tiao and Tsay apply a similar approach even to linear models), only that it is usually the accepted approach associated with neural networks, and less typical with conventional econometric linear or nonlinear models.

While it is conceivable that OMLR model construction could also follow the six-step procedure to make it more accommodating of alternative loss functions, given the choice of explanatory variables, it is not appealing to consider certain variations of the estimation algorithm involved. For example, nonconvergence in nonlinear estimation is usually not considered a virtue in model-building. Of course, this follows from the modeler's strong belief that the choice of functional form precluded overfitting in the first place. The BPN modeler, on the other hand, because of the infinite number of model configuration possibilities, is acutely aware of the possibility that overfitting may be reduced in two ways: by reducing the number of hidden weights to be estimated (akin to a parsimonious econometric model), or by estimating the weights less precisely (i.e., not iterating to convergence, or not overtraining, in neural network parlance). Explicit consideration of this latter possibility could add a dimension to economic model building not usually considered by applied economists. Furthermore, given that the true underlying economic data-generating model is rarely known, statistical analyses of out-of-sample predictions could proceed in the same fashion whether the predictions came from models which had converged or from models which had not (as is often done in the forecasting literature, when forecast RMSEs are compared between forecasts whose generating processes are totally unknown).

Conclusions

In out-of-sample prediction of Kansas farmers' responses to five surveyed questions involving risk, the ordered multinomial logistic regression model was shown to often perform statistically better than a naive model that uses no explanatory data and only randomly draws from in-sample known responses. However, the prediction superiority is dependent upon the choice of prediction accuracy test statistics. Specifically, the ordered logit model predicts well in an out-of-sample, root-mean-squared forecasting error sense. In this sense, it appears to predict well those response categories that are substantially represented in the data, with little allowance for a categorical response that may not be of interest to a researcher (such as the "unsure" response)—there is little allowance for alternatively specified loss functions.

On the other hand, feedforward backpropagation neural networks, coupled with a rigorous out-of-sample model testing procedure, are shown to accommodate alternatively specified loss functions. That is, the flexibility of the neural network model combined with the flexibility within the typically used estimation algorithm, backpropagation, allows for creation of predictive models that predict well those categorical responses that a researcher may be interested in. Predictive neural network models, chosen based on applying a test statistic of interest to out-of-sample predictions, are shown to perform well, based upon the same test statistic, when used to make still further out-of-sample predictions. The estimated neural network models suggest the possibility that nonconvergence of estimated parameters may be associated with enhanced out-of-sample prediction, a dimension of economic model-building which may have rewards, but certainly requires additional research.

Finally, ordered multinomial logit response models and neural networks both offer the potential for forecasting a farmer's self-assessment of risk. However, if a firm is specifically interested in those individuals who have strong perceptions, the neural network approach appears to outperform the ordered multinomial logit response models in segregating those individuals. Firms that market risk management products may be able to more efficiently design and implement sales strategies if they are able to better segregate potential customers. Neural networks offer a promising avenue to do just that.

[Received January 1995; final revision received January 1996.]

References

Amemiya, T. "Qualitative Response Models: A Survey." *J. Econ. Lit.* 19(December 1981):1483–1536.

Anderson, J.R., J.L. Dillon, and B. Hardaker. *Agricultural Decision Analysis.* Ames IA: Iowa State University Press, 1977.

Barkley, A.P., and B.L. Flinchbaugh. "Farm Operator Opinion and Agricultural Policy; Kansas Survey Results." *N. C. J. Agr. Econ.* 12(July 1990):223–39.

Dorsey, R., and W.J. Mayer. "Genetic Algorithms of Estimation Problems with Multiple Optima, Nondifferentiability, and Other Irregular Features." *J. Bus. and Econ. Statist.* 13(January 1995):53–66.

Fausett, L. *Fundamentals of Neural Networks, Ar-*

chitectures, Algorithms, and Applications. Englewood Cliffs NJ: Prentice Hall, 1994.

Goodwin, B.K., and T. Kastens. "Adverse Selection, Disaster Relief, and the Demand for Multiple Peril Crop Insurance." Research Report Prepared for the Federal Crop Insurance Corporation, Project No. 92-EXCA-3-0209, 17 May 1993.

Greene, W.H. *Econometric Analysis*, 2nd ed. New York: MacMillan, 1993.

Grudnitski, G., and L. Osburn. "Forecasting S&P and Gold Futures Prices: An Application of Neural Networks." *J. Futures Mkts.* 13(September 1993):631–43.

Hagan, M.T., and M.B. Menhaj. "Training Feedforward Networks with the Marquardt Algorithm." *IEEE Transactions on Neural Networks* 5(November 1994):989–93.

Hornik, K., M. Stinchcombe, and H. White. "Multilayer Feedforward Networks Are Universal Approximators." *Neural Networks* 2(1989):359–66.

Joerding, W.H., Y. Li, and D.L. Young. "Feedforward Neural Network Estimation of A Crop Yield Response Function." *J. Agr. and Appl. Econ.* 26(July 1994):252–63.

Kastens, T.L., A.M. Featherstone, and A.W. Biere. "A Neural Networks Primer for Agricultural Economists." *Agr. Finan. Rev.* 55(1995):54–73.

Kastens, T.L., and B.K. Goodwin. "An Analysis of Farmers' Policy Attitudes and Preferences for Free Trade." *J. Agr. and Appl. Econ.* 26(December 1994):497– 505.

Kosko, B. *Neural Networks and Fuzzy Systems.* Englewood Cliffs NJ: Prentice-Hall, 1992.

Matlab Numeric Computation Software. Natick MA: Math Works Inc., 1994.

Miller, L.H. "Neural Networks in Agricultural Credit Management." Paper presented at AAEA annual meeting, San Diego CA, 7–10 August 1994.

More, J.J. "The Levenberg-Marquardt Algorithm: Implementation and Theory." *Numerical Analysis.* G.A. Watson, ed. New York: Springer-Verlag, 1977.

NeuralWare. *Using NeuralWorks: A Tutorial for NeuralWorks Professional II/PLUS and Neural Works Explorer.* Pittsburgh PA, 1993.

Schurle, B., and W.I. Tierney. "A Comparison of Risk Preference Measurements With Implications for Extension Programming." Dept. Agr. Econ. Staff Paper No. 91-6, Kansas State University, July 1990.

Thomas, A.C. "Risk Attitudes Measured by the Internal Approach: A Case Study of Kansas Farmers." *Amer. J. Agr. Econ.* 69(December 1987):1101–5.

Tiao, G.C., and R.S. Tsay. "Some Advances in Nonlinear and Adaptive Modelling in Time-Series." *J. Forecasting* 13(March 1994):109–31.

Vandeveer, M.L., and E.T. Loehman. "Farmer Response to Modified Crop Insurance: A Case Study of Corn in Indiana." *Amer. J. Agr. Econ.* 76(February 1994):128–40.

Zavoina, R., and W. McElvey. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *J. Mathematical Sociology* 4(1975):103–20.