Forecasting Agricultural Commodity Prices Using Multivariate Bayesian Machine Learning Regression

by

Andres M. Ticlavilca, Dillon M. Feuz, and Mac McKee

Forecasting Agricultural Commodity Prices Using Multivariate Bayesian Machine

Learning Regression

Andres M. Ticlavilca,

Dillon M. Feuz,

and

Mac McKee [*]

---

[*] Andres M. Ticlavilca (andres.t@aggiemail.usu.edu) is a Graduate Research Assistant at the Utah Water Research Laboratory, Department of Civil and Environmental Engineering, Utah State University. Dillon M. Feuz (dillon.feuz@usu.edu) is a Professor at Utah State University in the Applied Economics Department. Mac McKee (mac.mckee@usu.edu) is the Director of the Utah Water Research Laboratory and Professor at the Department of Civil and Environmental Engineering, Utah State University.

Forecasting Agricultural Commodity Prices Using Multivariate Bayesian Machine Learning Regression

The purpose of this paper is to perform multiple predictions for agricultural commodity prices (one, two and three month periods ahead). In order to obtain multiple-time-ahead predictions, this paper applies the Multivariate Relevance Vector Machine (MVRVM) that is based on a Bayesian learning machine approach for regression. The performance of the MVRVM model is compared with the performance of another multiple output model such as Artificial Neural Network (ANN). Bootstrapping methodology is applied to analyze robustness of the MVRVM and ANN.

Keywords: Commodity prices, Forecasting, Machine learning, Bayesian

Introduction

The last few years there has been an increase in the volatility of many agricultural commodity prices. This has increased the risk faced by agricultural producers. The main purpose of agricultural commodity price forecasting is to allow producers to make better-informed decisions and to manage price risk.

Simple price forecast models such as naïve, or distributed-lag models have performed quite well in predicting agricultural commodity prices (Hudson, 2007). Other models such as "deferred future plus historical basis" models (Kastens et al., 1998), autoregressive integrated moving average (ARIMA) models and composite models (Tomek and Myers, 1993) lead to more accurate estimates. However, as the accuracy increases, so does the statistical complexity (Hudson, 2007).

Practical applications of more complex models are limited by the lack of required data and the expense of data acquisition. On the other hand, the increased volatility in agricultural commodity prices may increase the difficulty of forecasting accurately making the simple methods less reliable and even the more complex forecast methods may not be robust in this new market environment. To overcome these limitations, machine learning (ML) models can be used as an alternative to complex forecast models.

ML theory is related to pattern recognition and statistical inference wherein a model is capable of learning to improve its performance on the basis of its own prior experience (Mjolsness and DeCoste, 2001). Examples of ML models include the artificial neural networks (ANNs), support vector machines (SVMs) and relevance vector machines (RVMs).

ML models have been applied in financial economics modeling. Enke and Thaworwong (2005) used data mining and ANNs to forecast stock market returns. Co and Boosarawongse (2007) demonstrated that ANNs outperformed exponential smoothing and ARIMA models in forecasting rice exports. Also ML models have been applied in forecasting agricultural commodity prices. Shahwan and Odening (2007) used a hybrid between ANNs and ARIMA model to predict agricultural commodity prices.

The paper reported here presents a ML model to perform monthly multi-step-ahead predictions with predictive confidence intervals for agricultural commodity prices. Therefore, the model recognizes the patterns between multivariate outputs (future commodity prices) and multivariate inputs (past data collected about commodity prices).

This paper applies the Multivariate Relevance Vector Machine (MVRVM) (Thayananthan, 2005; Thayananthan et al., 2008). The MVRVM is an extension of the RVM algorithm developed by Tipping and Faul (2003). It can be used to produce multivariate outputs with confidence interval, via its Bayesian approach. The MVRVM has the same capabilities of the conventional RVM: high prediction accuracy, robustness, and characterization of uncertainty in the predictions.

The remainder of the paper describes the MVRVM approach, the model application for agricultural commodity price forecast, a comparison with an ANN model and conclusions.

Model Description

The MVRVM was developed by Thayananthan (2005) to provide an extension of the RVM algorithm for regression (Tipping, 2001; Tipping and Faul, 2003) to multivariate outputs.

Given a set of training examples of input-target vector pairs $\{\mathbf{x}^{(n)}, \mathbf{t}^{(n)}\}_{n=1}^{N}$, where N is the number of pattern observations, $\mathbf{x} \in R^D$ is a input vector, $\mathbf{t} \in R^M$ is a output target vector, the model learns the dependency between input and output target with the purpose of making accurate predictions of $\mathbf{t}$ for previous values of $\mathbf{x}$:

$$\mathbf{t} = \mathbf{W}\,\boldsymbol{\Phi}(\mathbf{x}) + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{W}$ is a M x P weight matrix and P = N+1. The error $\boldsymbol{\varepsilon}$ is assumed to be zero-mean Gaussian with diagonal covariance matrix $\mathbf{S}=\mathrm{diag}(\sigma_1^2, \ldots, \sigma_M^2)$.

$\boldsymbol{\Phi}(\mathbf{x})$ is a vector of basis functions of the form $\boldsymbol{\Phi}(\mathbf{x}) = [1, K(x,x^{(1)},\ldots K(x,x^{(N)}))$, where $K(\mathbf{x},\mathbf{x}_n)$ is a kernel function (Tipping, 2001, Thayananthan, 2008). In this paper, we considered a Gaussian kernel $K(x,x_n) = \exp(-r^{-2}\|x- x^{(n)}\|^2)$ where r is the kernel width parameter.

A likelihood distribution of the weights is defined as a product of Gaussians of the weight vectors ($\mathbf{w}_r$) corresponding to each output target ($\boldsymbol{\tau}_r$) (Thayananthan, 2008):

$$p(\{\mathbf{t}^{(n)}\}_{n=1}^{N} \mid \mathbf{W},\mathbf{S}) = \prod_{n=1}^{N} N(\mathbf{t}^{(n)} \mid \mathbf{W}\boldsymbol{\Phi}(\mathbf{x}^{(n)}),\mathbf{S}) = \prod_{r=1}^{M} N(\boldsymbol{\tau}_r \mid \mathbf{w}_r\boldsymbol{\Phi},\sigma_r^2) \tag{2}$$

where $\boldsymbol{\Phi} = [1, \boldsymbol{\Phi}(\mathbf{x}_1), \boldsymbol{\Phi}(\mathbf{x}_2),\ldots, \boldsymbol{\Phi}(\mathbf{x}_N)]$. To avoid overfitting from Equation (2), Tipping (2001) proposed constraining the selection of parameters by applying a Bayesian approach and defining an explicit zero-mean Gaussian prior probability distribution over the weights (Thayananthan, 2008):

$$p(\mathbf{W} \mid \mathbf{A}) = \prod_{r=1}^{M} \prod_{j=1}^{P} N(w_{rj} \mid 0, \alpha_j^{-2}) = \prod_{r=1}^{M} N(\mathbf{w}_r \mid 0, \mathbf{A}) \tag{3}$$

where $\mathbf{A} = \mathrm{diag}(\alpha_1^{-2}, \ldots, \alpha_P^{-2})^T$ is a diagonal matrix of hyperparameters α_j, and w_{rj} is the (r,j)th element of the weight matrix $\mathbf{W}$. Each α_j controls the strength of the prior over its associated weight (Tipping and Faul, 2003).

The posterior distribution of the weights is proportional to the product of the likelihood and prior distributions:

$$p(\mathbf{W} \mid \{\mathbf{t}\}_{n=1}^{N}, \mathbf{S}, \mathbf{A}) \propto p(\{\mathbf{t}\}_{n=1}^{N} \mid \mathbf{W}, \mathbf{S}) \, p(\mathbf{W} \mid \mathbf{A}) \tag{4}$$

Then, this posterior parameter distribution can be defined as the product of Gaussians for the weight vectors of each target (Thayananthan, 2008):

$$p(\mathbf{W} \mid \{\mathbf{t}\}_{n=1}^{N}, \mathbf{S}, \mathbf{A}) \propto p(\{\mathbf{t}\}_{n=1}^{N} \mid \mathbf{W}, \mathbf{S}) \, p(\mathbf{W} \mid \mathbf{A}) \propto \prod_{r=1}^{M} N(\mathbf{w}_r \mid \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \tag{5}$$

with covariance and mean, $\boldsymbol{\Sigma}_r = (\mathbf{A} + \sigma_r^{-2} \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1}$ and $\boldsymbol{\mu}_r = \sigma_r^{-2} \boldsymbol{\Sigma}_r \boldsymbol{\Phi}^T \boldsymbol{\tau}_r$, respectively.

An optimal weight matrix can be obtained by estimating a set of hyperparameters that maximizes the data likelihood over the weights in Equation (5) (Thayananthan, 2008). The marginal likelihood is then:

$$p(\{\mathbf{t}\}_{n=1}^{N} \mid \mathbf{A}, \mathbf{S}) = \int p(\{\mathbf{t}\}_{n=1}^{N} \mid \mathbf{W}, \mathbf{S}) \, p(\mathbf{W} \mid \mathbf{A}) d\mathbf{W} \,,$$
$$= \prod_{r=1}^{M} \int N(\boldsymbol{\tau}_r \mid \mathbf{w}_r \boldsymbol{\Phi}, \sigma_r^2) N(\mathbf{w} \mid 0, \mathbf{A}) = \prod_{r=1}^{M} \mid \mathbf{H}_r \mid^{-\frac{1}{2}} \exp(-\frac{1}{2} \boldsymbol{\tau}_r^T \mathbf{H}_r^{-1} \boldsymbol{\tau}_r) \tag{6}$$

where $\mathbf{H}_r = \sigma_r^2 \mathbf{I} + \boldsymbol{\Phi} \mathbf{A}^{-1} \boldsymbol{\Phi}^T$. The optimal set of hyperparameters $\boldsymbol{\alpha}^{opt} = \{\alpha_j^{opt}\}_{j=1}^{P}$ and noise parameters $(\boldsymbol{\sigma}^{opt})^2 = \{\sigma_r^{opt}\}_{r=1}^{M}$ are obtained by maximizing the marginal likelihood using the fast marginal likelihood maximization algorithm proposed by Tipping and Faul (2003). Many elements of $\boldsymbol{\alpha}$ go to infinity during the optimization process, for which the posterior probability of the weight becomes zero. These nonzero weights are called the relevance vectors (RVs) (Tipping and Faul, 2003).

Then, we can obtain the optimal covariance $\boldsymbol{\Sigma}^{opt} = \{\boldsymbol{\Sigma}_r^{opt}\}_{r=1}^{M}$ and mean $\boldsymbol{\mu}^{opt} = \{\boldsymbol{\mu}_r^{opt}\}_{r=1}^{M}$.

Given a new input $\mathbf{x}^*$, we can compute the predictive distribution for the corresponding output target $\mathbf{t}^*$ (Tipping, 2001) :

$$p(\mathbf{t}^* \mid \mathbf{t}, \boldsymbol{\alpha}^{opt}, (\boldsymbol{\sigma}^{opt})^2) = \int p(\mathbf{t}^* \mid \mathbf{W}, (\boldsymbol{\sigma}^{opt})^2) . p(\mathbf{W} \mid \mathbf{t}, \boldsymbol{\alpha}^{opt}, (\boldsymbol{\sigma}^{opt})^2) d\mathbf{W} \tag{7}$$

In Equation (7), both terms in the integrand are Gaussian. Then, this equation can be computed as:

$$p(\mathbf{t}^* \mid \mathbf{t}, \boldsymbol{\alpha}^{\text{opt}}, (\boldsymbol{\sigma}^{\text{opt}})^2) = N(\mathbf{t}^* \mid \mathbf{y}^*, (\boldsymbol{\sigma}^*)^2) \qquad (8)$$

where $\mathbf{y}^* = [\, y_1^*, ..., y_r^*, ... \, y_M^*]^T$ is the predictive mean with $y_r^* = (\boldsymbol{\mu}_r^{\text{opt}})^T \boldsymbol{\Phi}(\mathbf{x}^*)$; and $(\boldsymbol{\sigma}^*)^2 = [(\sigma_1^*)^2, ... (\sigma_r^*)^2, ..., (\sigma_M^*)^2]^T$ is the predictive variance with $(\sigma_r^*)^2 = (\sigma_r^{\text{opt}})^2 + \boldsymbol{\Phi}(\mathbf{x}^*)^T \boldsymbol{\Sigma}_r^{\text{opt}} \boldsymbol{\Phi}(\mathbf{x}^*)$ which contains the sum of two variance terms: the noise on the data and the uncertainty in the prediction of the weight parameters (Tipping, 2001).

The standard deviation σ_r^* of the predictive distribution is defined as a predictive error bar of y_r^* (Bishop 1995). Then, the width of the 90% Bayesian confidence interval for any y_r^* can be $\pm$ 1.65.σ_r^*. This Bayesian confidence interval (which is based on probabilistic approach) should not be confused with a classical frequentist confidence interval (which is based on the data).

Data

Monthly data for cattle, hog and corn prices were obtained for 21 years. The data were obtained from the Livestock Marketing Information Center website www.lmic.info and the data are initially collected and reported by the USDA-AMS. Corn prices were from Omaha, NE market; cattle prices were from Nebraska live fed cattle market, and hog prices were from the Iowa/southern Minnesota market. These are all large markets that are frequently used as standards by which to judge other markets.

Procedures

Monthly data from 1989 to 2003 were used to train each model and estimate the model parameters. Monthly data from 2004 to 2009 were used to test the models.

The inputs used in the model to predict monthly commodity price are expressed as

$$\mathbf{x} = [\mathbf{x}_{\text{tp-m}}]^T \qquad (9)$$

where,
tp = time of prediction
m = number of months previous to the prediction time
$\mathbf{x1}_{\text{tp-m}}$ = Commodity price 'm' months previous to the prediction time

The multiple output target vector of the model is expressed as

$$\mathbf{t} = [\, t_{\text{tp+1}}, t_{\text{tp+2}}, t_{\text{tp+3}}]^T \qquad (10)$$

where,
$t_{\text{tp+1}}$ = prediction of commodity price one month ahead
$t_{\text{tp+2}}$ = prediction of commodity price two months ahead
$t_{\text{tp+3}}$ = prediction of commodity price three months ahead

Performance evaluation

The kernel width is a smoothing parameter which defines a basis function to capture patterns in the data. This parameter cannot be estimated with the Bayesian approach. For this paper, a sensitivity analysis is performed to estimate the kernel width that gives accurate test results. The statistics used for the selection of the model is the root mean square percentage error (RMSPE), and is given by:

$$RMSPE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{t-t*}{t}\right)^2} .100 \qquad (11)$$

where t is the observed output; t* is the predicted output and N is the number of observations.

The sensitivity analysis was done by building several MVRVM models with variation in the kernel width (from 1 to 60) and the number of previous months required as input (from 1 to 12 months). The selected model was the one with the minimum RMSPE of the average outputs corresponding to the testing phase.

Bootstrap analysis

The bootstrap method (Efron and Tibshirani,1998) was used to guarantee robustness of the MVRVM (Khalil et al., 2005a). The bootstrap data set was created by randomly selecting from the whole training data set, with replacement. This selection process was independently repeated 1,000 times to yield 1,000 bootstrap training data sets, which are treated as independent sets (Duda et al., 2001). For each of the bootstrap training data sets, a model was built and evaluated over the original test data set.

A robust model is one that shows a narrow confidence bounds in the bootstrap histogram (Khalil et al., 2005b). A narrow confidence bounds implies low variability of the statistics with future changes in the nature of the input data, which indicates that the model is robust.

Comparison between MVRVM and ANN

A comparative analysis between the developed MVRVM and ANNs is performed in terms of performance and robustness. Readers interested in greater detail regarding ANN and its training functions are referred to Demuth et al. (2009).

Several feed-forward ANN models were trained and tested with variation in the type of training function, size of layer(from 1 to 10) and the number of months previous to the prediction time (from 1 to 12 months). The selected model was the one with the minimum RMSPE of the average outputs corresponding to the testing phase.

Results and Discussions

Table 1 shows the selected kernel width and the number of input months for each commodity price forecasting. The selected models were the ones with the lowest RMSPE of the average results.

Table 1 Selected MVRVM for each commodity (testing phase)

	Number of previous months	Kernel width	RMSPE (%)			Average RMSPE (%)
			1-month	2-months	3-months	
Corn	2	26	6.0	9.7	13.2	9.6
Cattle	8	14	3.1	5.0	6.3	4.8
Hog	11	32	6.1	7.8	9.6	7.8

Figures 1-3 show the predicted outputs (full lines) of the MVRVM for the testing phase for corn, cattle and hog prices respectively. These figures also show the 0.90 Bayesian confidence interval (shaded region) associated with the predictive variance $(\sigma_r*)^2$ of the MVRVM in Equation (8). We can see that this Bayesian confidence interval appears to be unchanged during the whole test period. As we mentioned in section 2, the predictive variance is $(\sigma_r*)^2 = (\sigma_r^{opt})^2 + \Phi(x*)^T \Sigma_r^{opt} \Phi(x*)$. The first term depends on the noise on the training data and the second term depends on the prediction of the parameter when a new input x* is given. From our results, it was found that there is a significant contribution from the first term (the noise variance on the training data) which made the contribution from the second term very small (close to zero). That is why the width of the confidence interval for the test results appears to be almost constant.

The model learns the patterns for one and two months ahead for the three commodities (Figures 1a, 1b, 2a, 2b, 3a and 3b). The performance accuracy is reduced for the three-month ahead price prediction of the three commodities (Figures 1c, 2c and 3c). This accuracy reduction is found in most of the multiple-time-ahead prediction models, where the farther we predict into the future, the less accurate the prediction becomes. Also, we can see that the model performance decreases in early 2008 for the corn price predictions. As we mentioned in section 4, monthly data from 1989 to 2003 were used to train the model and estimate the model parameters. The model probably needs to be retrained at this period (early 2008) and also it may need more number of previous months as inputs data during this period. However, we prefer not to provide more definite conclusions, as they might not be sufficiently well supported. More detailed analysis regarding strategies to improve model performance will be carried out for future research.
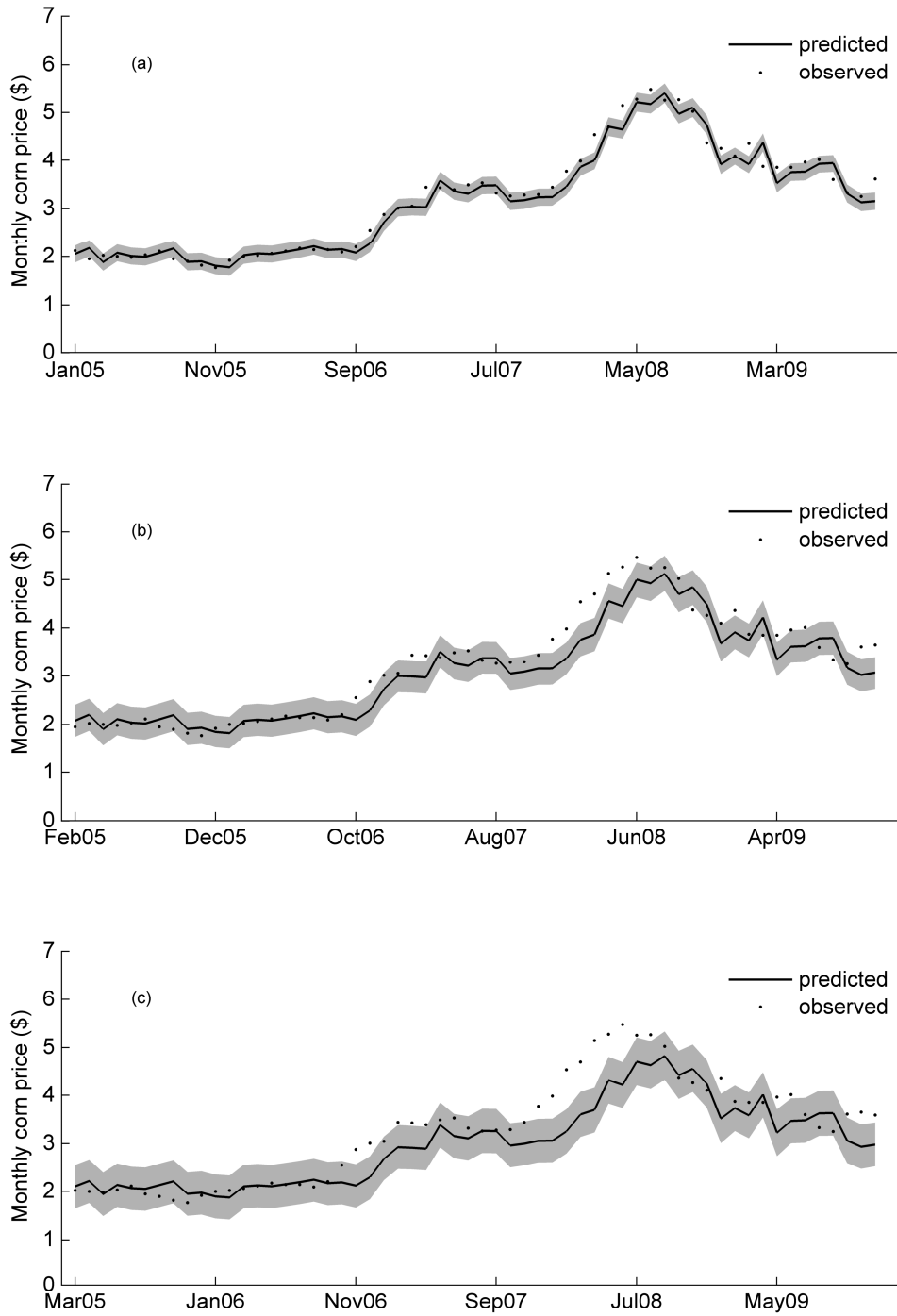
Figure 1. Observed versus predicted monthly corn price of the MVRVM model with 0.90 Bayesian confidence intervals (shaded region) for the testing phase: (a) 1-month ahead, (b) 2-months ahead, (c) 3-months ahead
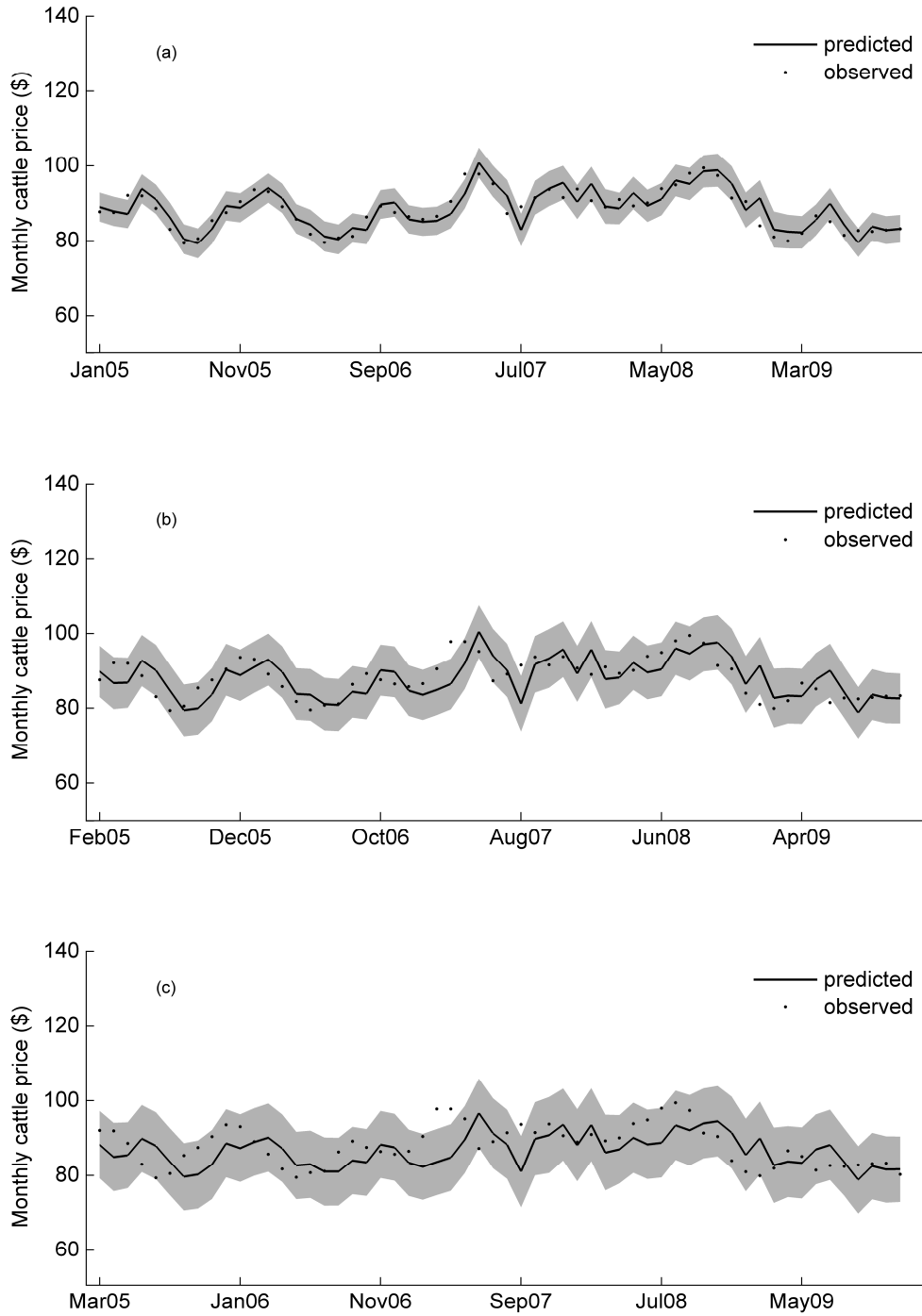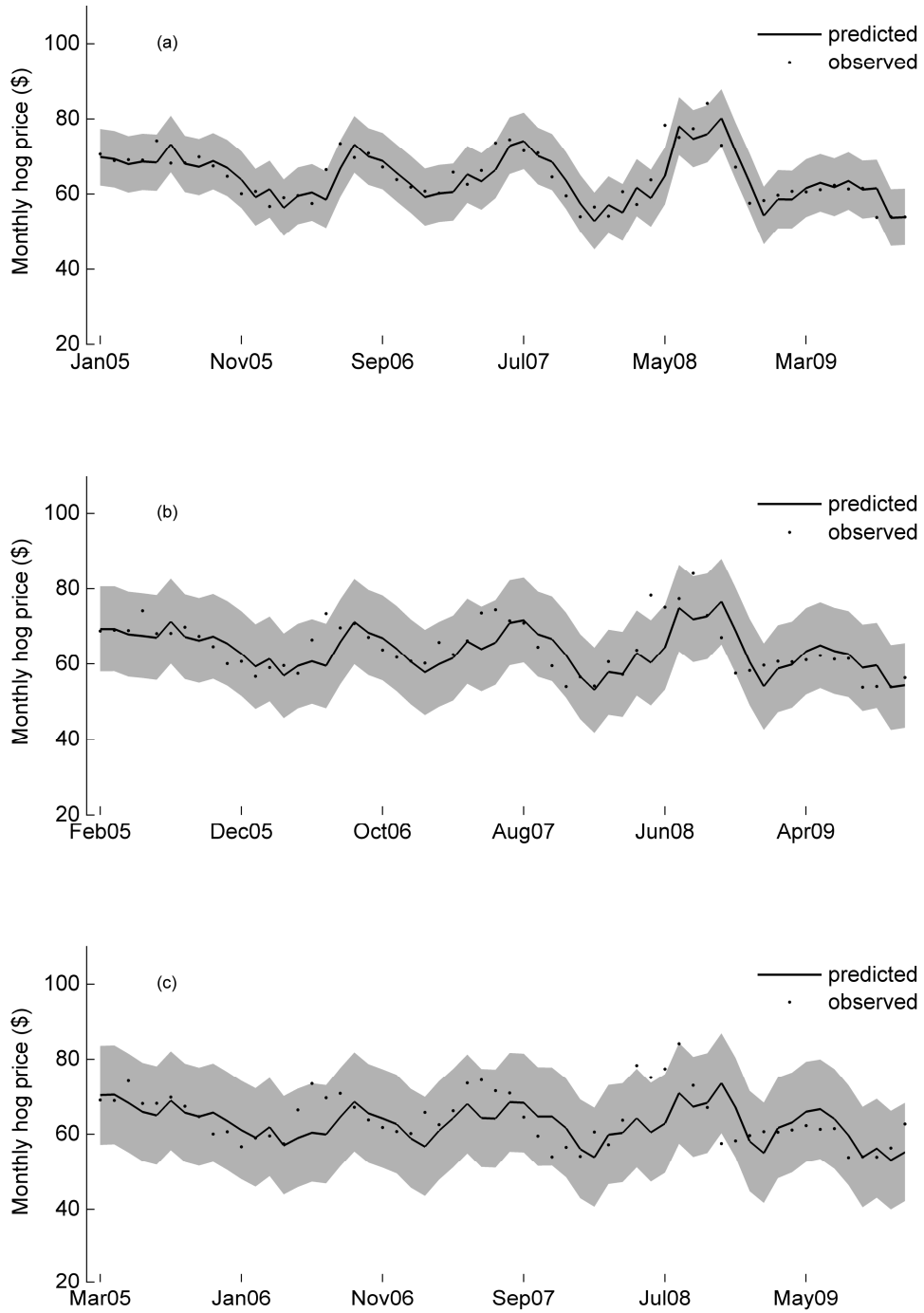
Figure 2. Observed versus predicted monthly cattle price of the MVRVM model with 0.90 Bayesian confidence intervals (shaded region) for the testing phase: (a) 1-month ahead, (b) 2-months ahead, (c) 3-months ahead

Figure 3. Observed versus predicted monthly hog price of the MVRVM model with 0.90 Bayesian confidence intervals (shaded region) for the testing phase: (a) 1-month ahead, (b) 2-months ahead, (c) 3-months ahead

Table 2 shows the selected ANN models for two types of training function. The model with conjugated-gradient-training function shows the lowest RMSPE of the average results for corn. The model with resilient-backpropagation-training function shows the lowest RMSPE of the average results for cattle and hog. Therefore they were selected as the best type of training function for each model that describes the input-output patterns.

Table 2. Selected ANN models for two types of training function (testing phase)

	RMSPE (%)									Average RMSPE (%)		
	Corn			Cattle			Hog					
Type of training function	1-month	2-months	3-months	1-month	2-months	3-months	1-month	2-months	3-months	Corn	Cattle	Hog
Resilient backpropagation	8.2	10.4	12.9	3.6	5.4	6.8	6.5	7.8	9.8	10.5	5.3	8.0
Conjugate gradient with Powell-Beale restarts	7.2	9.3	11.8	3.5	5.5	7.1	6.2	8.6	10.3	9.5	5.4	8.4

Table 3 shows the selected size of layer and the number of input months for each commodity price forecasting for the ANN models.

Table 3. Selected ANN for each commodity price prediction (testing phase).

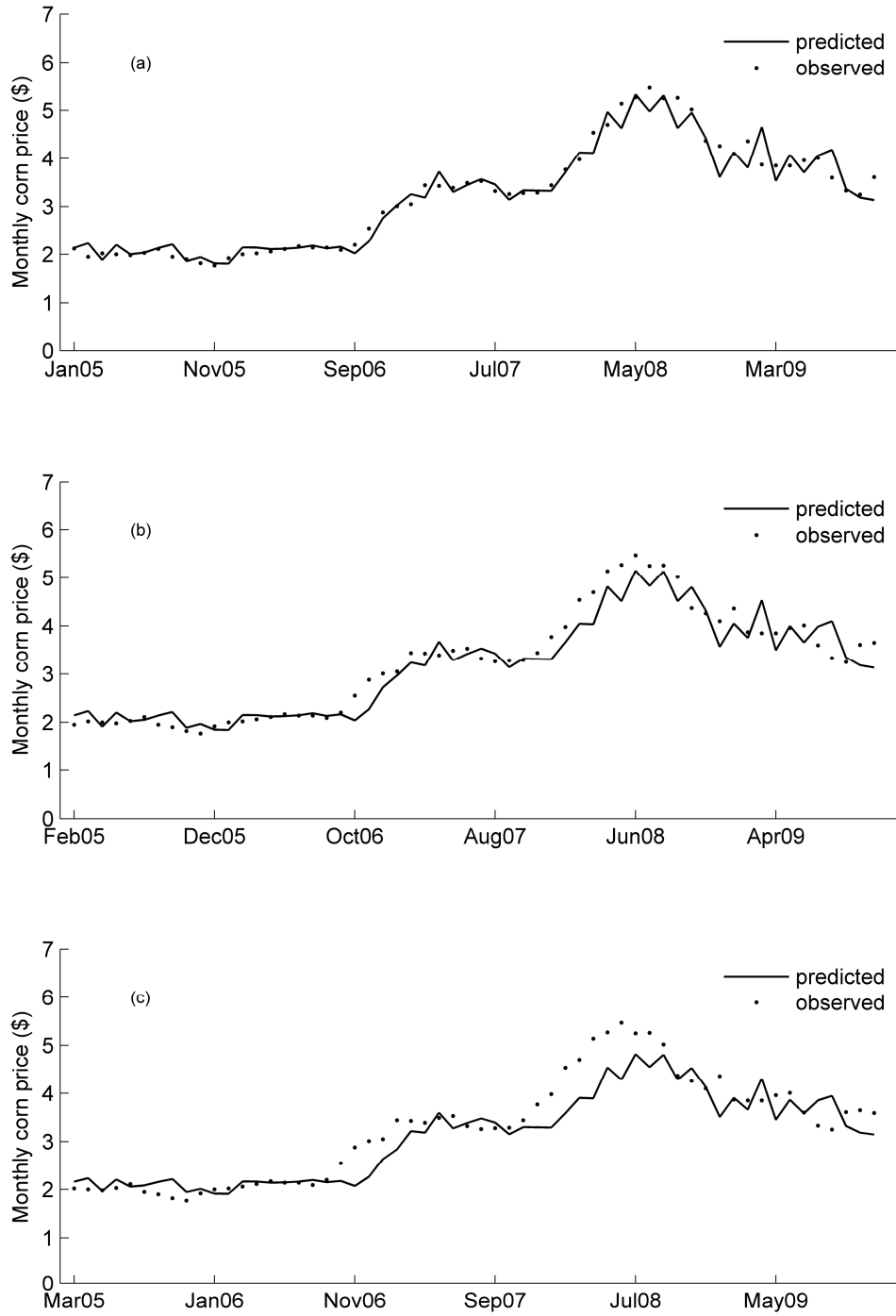| | Number of previous months | Size of layer | RMSPE (%) | | | Average RMSPE (%) |
			1-month	2-months	3-months	
Corn	10	2	7.2	9.3	11.8	9.5
Cattle	4	1	3.6	5.4	6.8	5.3
Hog	8	4	6.5	7.8	9.8	8.0

Figure 4. Observed versus predicted monthly corn price of the ANN model for the testing phase: (a) 1-month ahead, (b) 2-months ahead, (c) 3-months ahead
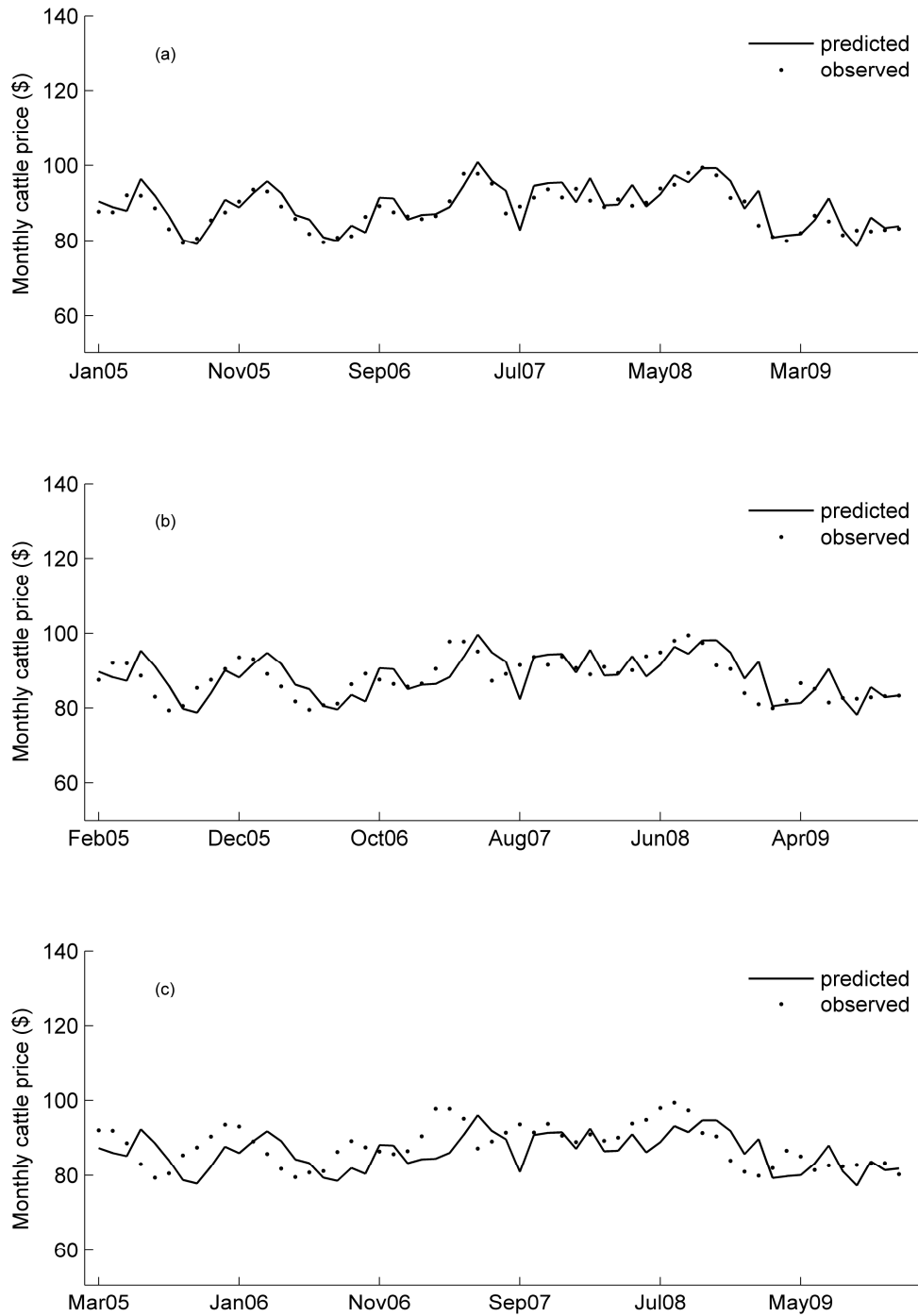
Figure 5. Observed versus predicted monthly cattle price of the ANN model for the testing phase: (a) 1-month ahead, (b) 2-months ahead, (c) 3-months ahead
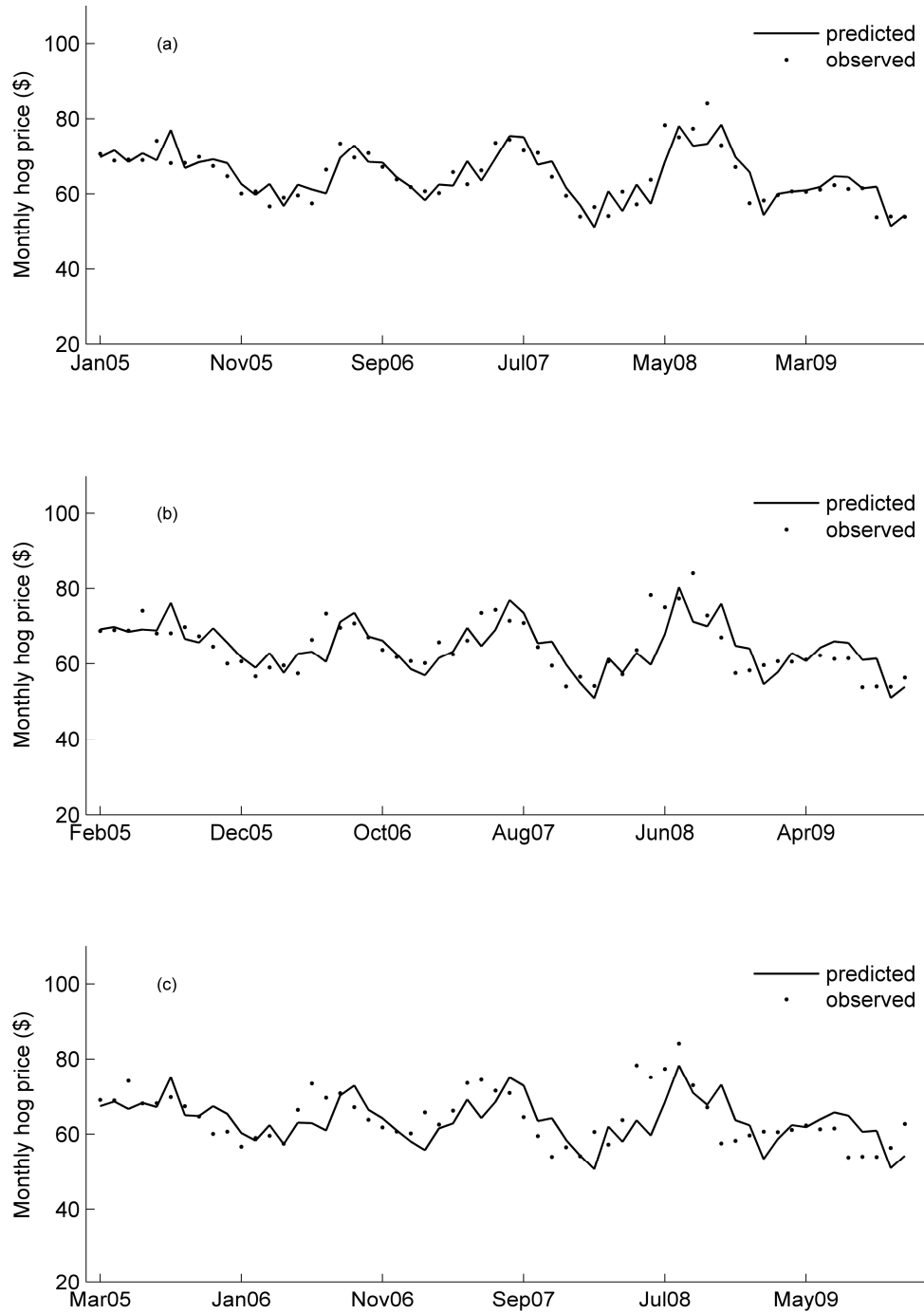
Figure 6. Observed versus predicted monthly hog price of the ANN model for the testing phase: (a) 1-month ahead, (b) 2-months ahead, (c) 3-months ahead

13

Figures 4-6 show the observed (dots) and predicted (full lines) outputs of the ANN for the testing phase for corn, cattle and hog prices respectively. RMSPE and RMSE statistics for the MVRVM and ANN prediction performance are displayed in Table 4. We can see that the MVRVM outperforms, has a smaller forecast error, the ANN for corn price prediction one month ahead, cattle price prediction one, two and three months ahead, and hog price prediction one and three months ahead. On the other hand, ANN outperforms MVRVM for corn price prediction two and three months ahead. Hog price predictions for two months ahead are similar for both models.

The ANN cattle model (Figure 5) appears to be shifted with a 1-3 month lag. This ANN model may need more number of previous months as inputs data during this period. On the other hand, the MVRVM cattle model (Figure 2) can overcome the performance lag problems since its Bayesian approach allows us to calculate predictive confidence intervals , instead of just providing a single target output (Bishop 1995) such as is the ANN model results.

Table 4. Model Performance using RMSPE and RMSE (testing phase)

Model	Statistics	Corn			Cattle			Hog		
		1-month	2-months	3-months	1-month	2-months	3-months	1-month	2-months	3-months
MVRVM	RMSPE (%)	6.0	9.7	13.2	3.1	5.0	6.3	6.1	7.8	9.6
	RMSE ($)	0.22	0.36	0.52	2.74	4.43	5.63	4.07	5.36	6.41
ANN	RMSPE (%)	7.2	9.3	11.8	3.6	5.4	6.8	6.5	7.8	9.8
	RMSE ($)	0.26	0.33	0.45	3.19	4.70	6.03	4.23	5.31	6.29

Figures 7-12 show the bootstrap histograms for the RMSE test based on 1,000 bootstrap training data sets of the MVRVM and ANN models for corn, cattle and hog prices respectively. The bootstrapped histograms of the MVRVM models (Figures 7, 9 and 11) show narrow confidence bounds in comparison to the histograms of the ANN models (Figures 8, 10 and 12). Therefore, the MVRVM is more robust.
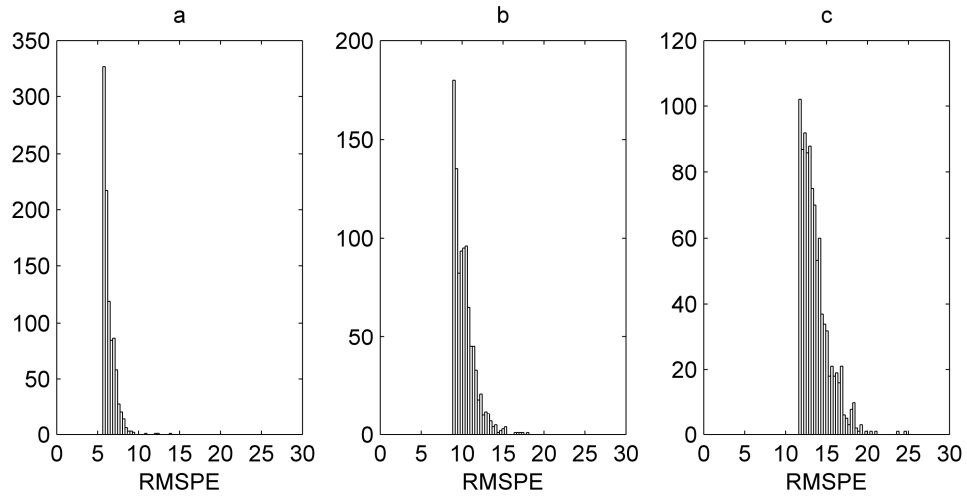
Figure 7. Bootstrap histograms of the MVRVM model of corn price predictions for the RMSPE: a) one month ahead, b) two months ahead, c) three months ahead
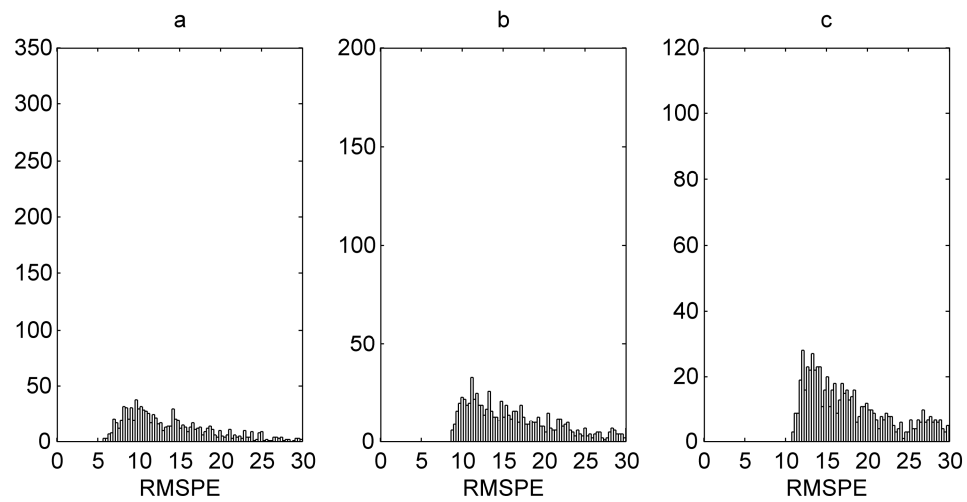

Figure 8. Bootstrap histograms of the ANN model of corn price predictions for the RMSPE: a) one month ahead, b) two months ahead, c) three months ahead
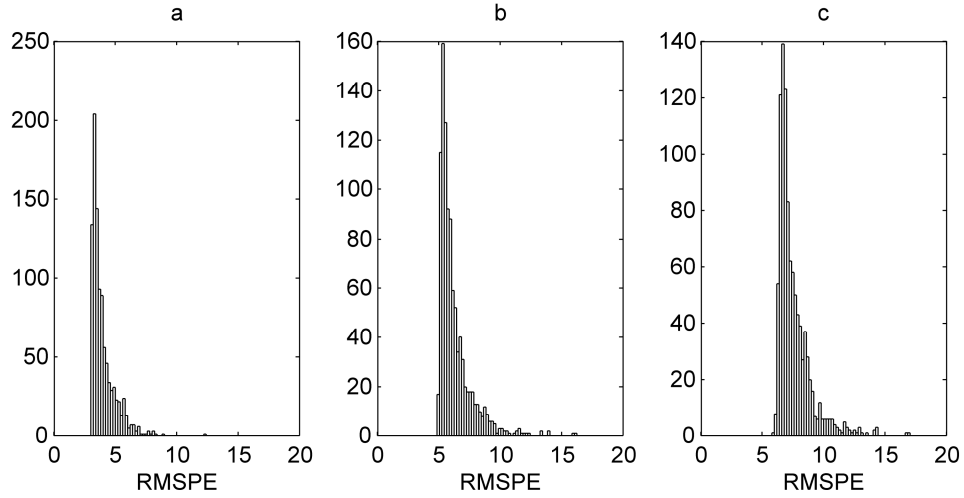
Figure 9. Bootstrap histograms of the MVRVM model of cattle price predictions for the RMSPE: a) one month ahead, b) two months ahead, c) three months ahead
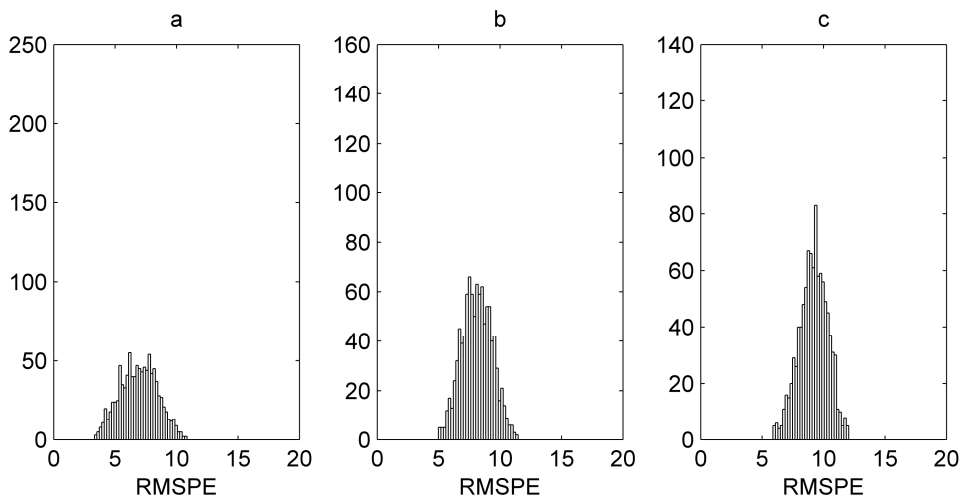

Figure 10. Bootstrap histograms of the ANN model of cattle price predictions for the RMSPE: a) one month ahead, b) two months ahead, c) three months ahead
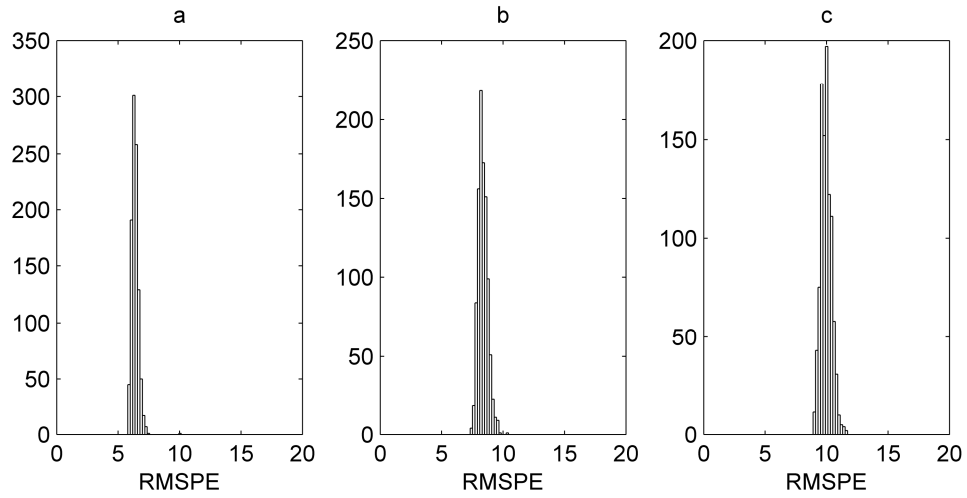
Figure 11. Bootstrap histograms of the MVRVM model of hog price predictions for the RMSPE: a) one month ahead, b) two months ahead, c) three months ahead
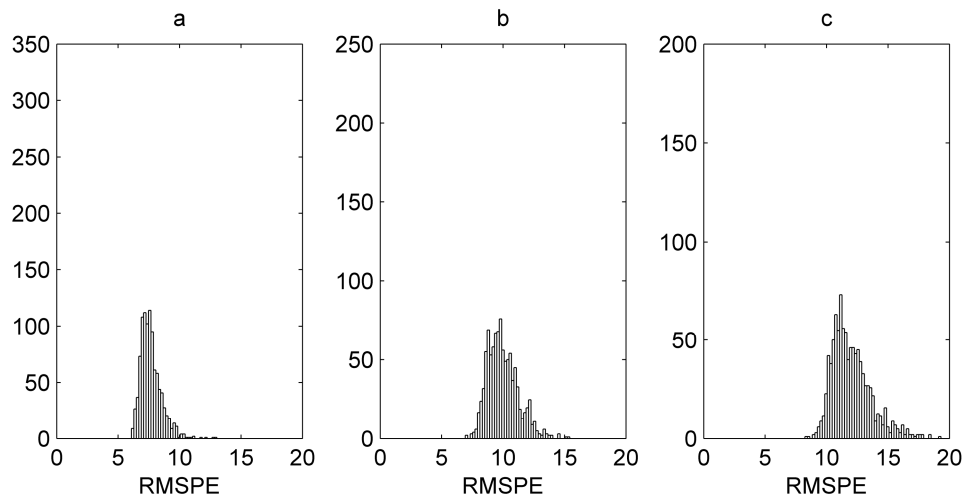

Figure 12. Bootstrap histograms of the ANN model of hog price predictions for the RMSPE: a) one month ahead, b) two months ahead, c) three months ahead

Summary and Conclusions

This paper applies a MVRVM model to develop multiple-time-ahead predictions with confidence intervals of monthly agricultural commodity prices. The predictions are one, two and three months ahead of prices of cattle, hogs and corn. The MVRVM is a regression tool extension of the RVM model to produce multivariate outputs.

The statistical test results indicate an overall good performance of the model for one and two month's prediction for all the commodity prices. The model performance decreased in early 2008 for the corn price predictions. The performance also decreased for the three-month prediction of the three commodity prices.

The MVRVM model outperforms the ANN most of the time with the exception of corn price prediction two and three months ahead. However, the bootstrap histograms of the MVRVM model show narrow confidence bounds in comparison to the histograms of the ANN model for the three commodity price forecasts. Therefore, the MVRVM is more robust.

The results presented in this paper have demonstrated the overall good performance and robustness of MVRVM for simultaneous multiple-time-ahead predictions of agricultural commodity prices . The potential benefit of these predictions lies in assisting producers in making better-informed decisions and managing price risk.

Future work

In this research, we have not analyzed the sparse property (low complexity) of the MVRVM since we have worked with relatively small data set (166 monthly observations) to train the model. Future research will be performed by analyzing weekly price (more than 1000 observations) and fully exploit the sparse characteristics of the Bayesian approach when dealing with large dataset.

Also, the relevance vectors (RVs) (which are related to the sparse property) are the summary of the most essential observations of the training data set to build the MVRVM. In this paper we have not analyzed the RVs since we are dealing with small data set. Future research will be performed by analyzing with more details the statistical meaning of RVs with respect to a large training data set. For example, this analysis can also be related to whether we would recommend reducing the number of historical data observations for retraining the model with new data in the future.

The kernel width and the number of previous months required as input cannot be estimated with the Bayesian approach. For this paper, a sensitivity analysis (by trial and error) was performed to estimate these parameters that gave accurate test results. We could see that the overall test results are good. However, the model performance decreases for some periods (i.e. early 2008 for the corn price predictions). Therefore, more analysis regarding the selection of these parameters (kernel width an number of previous months as inputs) will be carried out in a follow-up paper.

Application of a hybrid model (e.g. Bayesian approach embedded in ANN model) will be applied and compared to the MVRVM model in terms of accuracy, complexity and robustness.

References

Bishop, C. M. Neural Networks for Pattern Recognition (1995). Oxford University Press UK.

Co, H. C., and R. Boosarawongse (2007). "Forecasting Thailand's Rice Export: Statistical Techniques vs. Artificial Neural Networks." *Computers and Industrial Engineering* 53: 610-627.

Demuth, H., M. Beale, and M. Hagan (2009) Neural network toolbox user's guide, The MathWorks Inc, MA, USA.

Duda, R. O., P. Hart and D. Stork (2001). Pattern Classification, edited by Wiley Interscience, Second Edition, New York.

Efron, B., and R. Tibshirani (1998) An introduction of the Bootstrap. Monographs on Statistics and Applied Probability 57, CRC Press LLC, USA.

Enke, D. and S. Thawornwong (2005). The Use of Data Mining and Neural Networks for Forecasting Stock Market Returns. Expert Systems with Applications 29:927-940.

Hudson, D. (2007). Agricultural Markets and Prices, Maden, MA. Blackwell Publishing.

Kastens, T. L., R. Jones and T. C. Schroeder (1998). Future-Based Price Forecast for Agricultural Producers and Businesses. Journal of Agicultural and Resource Economics 23(1): 294-307.

Khalil, A., M. McKee, M. W. Kemblowski, and T. Asefa (2005a). Sparse Bayesian learning machine for real-time management of reservoir releases. Water Resources Research, 41, W11401, doi: 10.1029/2004WR003891.

Khalil, A., M. McKee, M. W. Kemblowski, T. Asefa, and L. Bastidas (2005b). Multiobjective analysis of chaotic dynamic systems with sparse learning machines, Advances in Water Resources, 29, 72-88, doi:10.1016/j.advwatres.2005.05.011.

Mjolsness, E., and D.DeCoste (2001). Machine learning for science: state of the art and future prospects. Science 293:2051–2055.

Shahwan, T and M Odening (2007). Forecasting Agricultural Commodity Prices using Hybrid Neural Networks. In Computational Intelligence in Economics and Finance. Springer, Berlin. 63-74.

Thayananthan, A. (2005). Template-based Pose Estimation and Tracking of 3D Hand Motion. PhD Thesis, Department of Engineering, University of Cambridge, Cambridge, United Kingdom.

Thayananthan, A., R. Navaratnam, B. Stenger, P.H.S. Torr, and R. Cipolla (2008). Pose estimation and tracking using multivariate regression, Pattern Recognition Letters 29(9), pp. 1302-1310.

Tipping, M. E. (2001), Sparse Bayesian learning and the relevance vector machine, Journal of Machine Learning, 1, 211–244.

Tipping, M., and A. Faul (2003). Fast marginal likelihood maximization for sparse Bayesian models, paper presented at Ninth International Workshop on Artificial Intelligence and Statistics, Soc. for Artif Intel and Stat, Key West, Florida.

Tomek, W. G., and R. J. Myers (1993). Empirical Analysis of Agricultural Commodity Prices: A Viewpoint, Review of Agricultural Economics, 15(1):181-202.