NCCC-134
APPLIED COMMODITY PRICE ANALYSIS, FORECASTING AND MARKET RISK MANAGEMENT

Accuracy-Informativeness Tradeoff for Interval Forecast Comparison

by

Olga Isengildina and Fabio Mattos

Accuracy-Informativeness Tradeoff
for Interval Forecast Comparison

Olga Isengildina,

and

Fabio Mattos [*]

Paper presented at the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management
St. Louis, Missouri, April 20-21, 2015

[*] Olga Isengildina (isengild@uta.edu) is an Associate Clinical Professor in the Department of Economics, University of Texas-Arlington, and Fabio Mattos (fmattos@unl.edu) is an Assistant Professor in the Department of Agricultural Economics, University of Nebraska-Lincoln.

Accuracy-Informativeness Tradeoff
for Interval Forecast Comparison

Price interval forecasts are analyzed in this study focusing on three main characteristics: coverage, error and informativeness. The tradeoff between accuracy and informativeness results from the fact that greater accuracy is achieved at the cost of lower informativeness and vice versa. The purpose of this paper is to evaluate user preferences for these characteristics using experimental methods. Contingent valuation methods were used to elicit user willingness to pay for forecasts with various characteristics. Estimation results demonstrate that coverage is the most important characteristic, followed by width and normalized error.

Keywords: coverage, error, informativeness, interval forecasts, forecast comparison, accuracy-informativeness tradeoff, precision.

Introduction

The need for probability and interval forecasting has been repeatedly expressed in the agricultural economics literature (e.g., Teigen and Bell, 1978; Timm, 1966; Bessler and Kling, 1989; Bessler, 1984). It has been long argued by academics (e.g., Armstrong, 2001) that prediction intervals are superior to point forecasts as they communicate information about uncertainty associated with the forecasts, which should aid in decision making. Onkal and Bolger (2004) argue that both forecast providers and forecast users prefer prediction intervals to point forecasts. However, application of interval and probability forecasts is still relatively scarce. Klassen and Flores (2001) and Dalrymple (1987) report that most companies use point forecasts rather than prediction intervals in their operations. Interval forecasts in agricultural economics are largely limited to commodity price forecasting. For example, United States Department of Agriculture (USDA) provides interval price forecasts for major field crops in their World Agricultural Supply and Demand Estimates (WASDE) reports. USDA's livestock price forecasts are published in an interval form in Livestock, Dairy and Poultry Outlook reports. Interval price forecasts for hogs are also available from Iowa State University (ISU). While these agricultural price forecasts are published in a form of a range, they do not specify the probability of the final value being contained within the interval (confidence level). This characteristic provides particular challenges for analysis of these interval forecasts as typical measures of interval forecast evaluation (Christoffersen, 1998) focus on calibration, which describes the difference between observed and stated confidence levels.

To the best of our knowledge, only three previous studies evaluated USDA price forecasts as intervals rather than reducing them to a point estimate. Sanders and Manfredo (2003) examined one-quarter-ahead WASDE interval forecasts of livestock prices from 1982 to 2002. Evaluation of hit rates, the proportion of time actual market prices fall in the forecasted ranges, revealed relatively low hit levels for livestock price forecasts, about 48% of the time for broilers, 41% of the time for cattle, and only 35% of the time for hogs. The authors did not conduct any formal tests of interval forecast accuracy other than showing that based on z-scores for testing equality in the proportion of hits, these forecasts were not significantly better than a proposed naïve alternative.

Isengildina, Irwin, and Good (2004) showed that monthly WASDE interval forecasts of corn and soybean prices during the 1980/81 through 2001/02 marketing years also had relatively low hit rates ranging from 36 to 82% for corn and from 59 to 89% for soybeans depending on the forecast month. The authors applied unconditional and conditional tests of interval forecast accuracy developed by Christoffersen (1998) to test whether WASDE price forecast intervals were calibrated at two benchmark confidence levels.

Isengildina and Sharp (2012) evaluated the implications of asymmetry on accuracy of USDA interval forecasts of corn, soybean and wheat prices. Although forecast intervals published by the USDA for corn, soybean and wheat prices are reportedly symmetric, they have shown that these intervals should not always be interpreted as symmetric. Their findings demonstrate that due to the uneven distribution of forecast misses around the interval, calibration of several corn, soybean and wheat price forecasts over 1980/81 through 2009/10 marketing years was rejected by basic coverage tests (suitable for symmetric intervals) but not rejected by the tests adjusted for asymmetry. In other words, these forecasts were asymmetric but accurate.

All of these previous analyses of agricultural interval forecasts have focused on a single characteristic, coverage, which is a binary measure of accuracy that reflects whether or not the interval includes the final value. An alternative measure, which describes the distance between the final value and the interval, similar to forecast error, can be viewed as a continuous measure of interval forecast accuracy (Yaniv and Foster, 1995, 1997). Furthermore, the true “value” of the forecast may contain several additional dimensions. For example, Fishhoff (1994) investigated why people sometimes pay scan attention to forecasts expressed in probabilistic terms. One reason was the perceived irrelevance of the forecasts. In contrast to point forecasts, wide prediction intervals may be seen as so lacking in definitiveness that they are seen as irrelevant and uninformative (Rush and Page, 1979; Yaniv and Foster, 1995, 1997) particularly by decision makers who have an intolerance of ambiguity. Thus, interval forecast informativeness is an important characteristic that should be taken into account. The interaction between the above characteristics is based on the tradeoff between accuracy and informativeness: greater accuracy is achieved at the cost of lower informativeness and vice versa.

Yaniv and Foster (1995) proposed the following measure of accuracy-informativeness tradeoff in their 1995 paper:

$$L = \frac{|t-m|}{g} + \alpha \ln(g), \tag{1}$$

where t is the final value; m is the interval midpoint; g is the width of the interval and α is a tradeoff parameter. The proposed model was hypothesized to predict students’ preferences for forecast intervals and was shown to perform better than alternative approaches. However, the model was selected subjectively and not based on empirical data.

The goal of this study is to develop a better understanding of user preferences for interval forecast coverage, error and informativeness using experimental techniques. Contingent valuation methods will be used to elicit user willingness to pay for forecasts with various characteristics. Generated data will allow evaluation of the relative importance of various characteristics in the user preference function.

The findings of this study will provide a comprehensive tool for selecting a preferred interval forecast. Our findings will be most relevant for interval forecast providers who are faced with tradeoffs between accuracy, precision and informativeness when they generate a forecast. Better understanding of consumer preferences for these characteristics would allow them to provide forecasts that consumers are looking for, which in turn may help consumers make better decisions.

Interval Forecast Characteristics

Three interval forecast characteristics are included in this analysis: coverage, error and informativeness. Coverage is a binary measure of interval forecast accuracy, which describes whether the forecast interval contains the final or "true" value (y_t) (also referred to as hit rate). This measure can be further used in forecast calibration tests (Christoffersen, 1998) to assess whether the observed coverage level is statistically different from the stated coverage level. The drawback of this measure is its binary, all-or-nothing nature which implies that forecast misses have no value.

Error describes a continuous measure of interval forecast accuracy used in this study. Yaniv and Foster (1995) proposed using a normalized error, an error-to-precision ratio, $\frac{|t-m|}{g}$, where t is the final value; m is the interval midpoint; and g is the width of the interval. Thus, the numerator is similar to a traditional measure of absolute error for point forecast. Division or "normalization" of error by interval width allows judging the forecast error in the context of how specific the forecast is, a \$10 error associated with a \$5 interval is judged differently from the \$10 error for a \$1 interval. Thus, this measure captures the interaction between error and precision. Measuring the error from the midpoint also provides insight for other characteristics: if the ratio is greater than 0.5, the forecast is inaccurate, if the ratio is 0, the forecast is symmetric and accurate.

Informativeness refers to how specific the intervals are based on their width. Following Yaniv and Foster (1995), informativeness is measured as ln(g), where g is the width of the interval. A logarithmic transformation of the interval width is consistent with the concave nature of human responses to changes in objective magnitudes.

As mentioned in the introduction, Yaniv and Foster (1995) proposed an additive measure of accuracy-informativeness tradeoff in their 1995 paper (equation 1). We will estimate their model using the data generated from our experiments and test the following additional hypotheses:

1) User preferences may be discontinuous for accurate vs inaccurate forecasts, including information about forecast calibration may serve as a shifter in the preference function;

2) User preferences may be asymmetric for positive and negative forecast errors.

3) Alternative functional forms may better reflect user preferences for forecast characteristics.

Choice Experiments

Forecast user preferences were measured in this study using willingness to pay (WTP) values elicited from the choice experiments. The participants were presented with the following context:

> *You have recently purchased 10 shares of stock of company X for $45 per share. Now you need to make decisions regarding whether to hold or sell these shares based on your expectations for future price movements. You can use professional forecasts of what would be your stock's price in 6 months to help you make these decisions. Due to high volatility of your stock's price, market analysts often present their forecasts as a range inside which the price is likely to fall.*

The initial forecast with $4 width and $50 stock price was priced at $30 and the respondents had to choose between alternative forecasts based on their characteristics and relative prices (Figure 1). Based on pairwise choices illustrated in Figure 1, the switching point in preferences was used to infer the WTP value. For example, for a set of choices given in Table 1, user preference switches from forecast A to forecast B when the price of forecast B declines from $38 to $36 and the WTP for forecast B is calculated as an average between two prices at which the switch occurs, or $37.

The questionnaire consisted of 3 parts: in the first part the stock price was held constant while the interval width varied. In the second part, the interval width was held constant while the stock price varied. Since in the third part both the width and the stock price varied, the reference price could not be determined and the respondents were asked directly what would be the maximum price they would have paid for the forecasts.

The experiments were conducted with university students majoring in economics and agricultural economics at the University of Texas at Arlington and the University of Nebraska-Lincoln in March-April 2015. A total of 39 completed questionnaires were used for the analysis. The next section presents the descriptive statistics of the data generated through choice experiments.

Data

In this section we examine differences in WTP data generated in choice experiments across the three interval forecast characteristics included in this analysis: coverage, error and informativeness.

Table 2 demonstrates that average WTP was \$6.46 higher for accurate forecasts (hit = 1). Similarly, average WTP was the highest (\$33.8) for forecasts with error-to-precision ratio (ETP) = 0 (accurate and symmetric) and declined initially as ETP increased in positive or negative direction. However, the decline in WTP with ETP was non-linear and inconsistent, as WTP was higher for large positive (3.5) and negative (-2.5; -6.5) ETP. A relatively small number of observations and interaction with other characteristics may potentially explain this inconsistency. A more consistent pattern was observed in changes in WTP with interval width, which was the highest for narrow, more informative forecasts and decreased as width increased (with exception of width = 2). Further investigation of how WTP is affected by forecast characteristics is conducted in the next section using regression analysis.

Results

Evaluation of the relative importance of various characteristics in the user preference function was conducted using a panel regression approach. Individual fixed effects are adopted to account for unobservable individual characteristics that vary across participants. The model is tested for the presence of heteroskedasticity and serial correlation in the residuals with a Breusch-Pagan test and the Breusch-Godfrey/Wooldridge test, respectively. Both null hypotheses that heteroskedasticity and serial correlation are not present can be rejected at the 1% level, thus panel-corrected standard errors (PCSE) are adopted.

The first step was to estimate Yaniv and Foster’s (1995) model using data generated from choice experiments. Since the coefficient of error-to-precision is assumed to be unity in equation (1), error variable was transferred to the left hand side and subtracted from the WTP level. Table 3 demonstrates that the tradeoff coefficient α estimated in this study to be 1.46 is about twice the size of the one estimated in Yaniv and Foster’s (1995) model, 0.74. The negative coefficient of the tradeoff parameter illustrates the opposite scales of the WTP and the L score in the Yaniv and Foster’s (1995) model. The constant adjusts the WTP value for the \$30 starting point used in our study. Yaniv and Foster’s model explains about 21% of the variation in the WTP values elicited from the choice experiments in this study.

The second model shown in Table 3 relaxes the assumption of unity coefficient for error-to-precision variable and introduces coverage as a separate variable to test for discontinuity in accuracy preferences. Our findings demonstrate that forecast users are willing to pay a premium of about \$4 for forecasts that contain the final value and thus their preferences for accuracy are not continuous. The coefficient on the absolute error to precision (ETP) ratio of -0.67 is smaller than 1 assumed in the Yaniv and Foster’s study, suggesting a less than proportional impact. The sign of the coefficient of ETP indicates an inverse relationship between normalized (relative) error and willingness to pay, which is consistent with expectations. In Model 3 we separate positive and negative errors and test for linearity in informativeness variable. Our results suggest that while negative errors maintain a predicted inverse relationship with WTP, positive errors exhibit a direct relationship with WTP. This pattern was also observed in the data and may be due to the lack of the observations with positive errors. Linear relationship between informativeness and WTP was not rejected further improving explanatory power of our model. Other functional forms (quadratic) were also explored but did not yield superior results.

Summary and Conclusions

This study sought to develop a better understanding of user preferences for interval forecast coverage, error and informativeness using experimental techniques. Contingent valuation methods were used to elicit user willingness to pay for forecasts with various characteristics. Panel regression analysis of these data revealed that user preferences for accuracy are non-linear and users are willing to pay a $4 premium for forecasts that contain a final value relative to less accurate forecasts. As forecast error grows, WTP decreases by about 67 cents every time the error grows by interval width, implying the WTP of zero for forecasts with the error of about 6 times the interval width. Wider intervals are also less valuable for forecast users, as our findings suggest that WTP decreases by about $1.19 for each 100% increase in interval width, or about $2.46 for each extra dollar of width. In relative terms, the magnitudes of the coefficients suggest that accuracy is the most important characteristic, followed by width and normalized error.

The specific values in the above findings should be interpreted with caution as they are associated with the reference point of $30 for an accurate $4 wide forecast used in this study. Further research that explores alternative reference points and forecast characteristics will allow achieving more general results.

References

Armstrong, J.S., 2001. *Principles of Forecasting*. Kluwer Academic Publishers, Boston.

Bessler, D. A. and J.L. Kling. "The Forecast and Policy Analysis." *American Journal of Agricultural Economics*, 71(1989):503-506.

Bessler, D.A. "Subjective Probability." in P.J. Barry (ed.) *Risk Management in Agriculture*. Iowa State University Press: Ames, Iowa, 1984; pp. 43-52.

Christoffersen, P.F. 1998."Evaluating interval forecasts." *International Economic Review*, 39: 841–862.

Dalrymple, D.J., 1987. Sales forecasting practices, results from a United States survey. *International Journal of Forecasting* 3, 379-391.

Fishhoff, B., 1994. What forecasts (seem to) mean. *International Journal of Forecasting* 10, 387-403.

Isengildina, O. and J. Sharp. 2012. "Evaluation of USDA Interval Forecasts Revisited: Asymmetry and Accuracy of Corn, Soybean, and Wheat Prices" *Agribusiness: An International Journal*, 28 (3): 310–323.

Isengildina, O., S.H. Irwin, and D. L. Good. 2004."Evaluation of USDA interval forecasts of corn and soybean prices." *American Journal of Agricultural Economics*, 86(4): 990–1004.

Klassen, R.D., and B.E. Flores, 2001. Forecasting practices of Canadian firms: Survey results and comparisons. *International Journal of Production Economics* 70, 163-174.

Onkal, D., and F. Bolger, 2004. Provider-user differences in perceived usefulness of forecasting formats. *Omega International Journal of Management Science* 32, 31-39.

Rush, H., and W. Page, 1979. Long-term metals forecasting: The track record 1910-1964. *Futures* 1, 321-337.

Sanders, D. R. and M. R. Manfredo. 2003. "USDA livestock price forecasts: A comprehensive evaluation." *Journal of Agricultural and Resource Economics*, 28(2): 316–334.

Teigen, L.D. and T.M. Bell. "Confidence Intervals for Corn Price and Utilization Forecasts." *Agricultural Economic Research*, 30(1978):23-29.

Timm, T.R. "Proposals for Improvement of the Agricultural Outlook Program in the United States." *Journal of Farm Economics*, 48(1966):1179-1184.

Yaniv, I. and D.P. Foster, “Graininess of Judgment Under Uncertainty: An Accuracy-Informativeness Trade-Off.” *Journal of Experimental Psychology: General*. 124(1995):424-432.

Yaniv, I. and D.P. Foster, “Precision and Accuracy of Judgmental Estimation.” *Journal of Behavioral Decision Making*, 10(1997):21-32.

Table1. Determination of WTP from Choice Experiments.

Forecast interval A [48 ; 52]	Prefer A	Prefer B	Forecast interval B [49 ; 51]
price			price
$30	X		$42
$30	X		$40
$30	X		$38
$30		X	$36
$30		X	$34
$30		X	$32

Table 2. Changes in WTP across Forecast Characteristics.

Forecast characteristic	Mean	Std. Dev.	Min.	Max.
		----$----		
Miss = 0	26.86	10.63	0.00	51.00
Hit = 1	33.32	7.53	0.00	51.00
ETP = -6.5	24.92	11.87	0.00	42.00
ETP = -3.0	21.00	5.59	10.00	29.50
ETP = -2.5	25.77	12.31	0.00	40.00
ETP = -2.0	24.95	4.01	19.00	33.50
ETP = -1.5	22.85	9.13	0.00	44.00
ETP = -1.0	34.41	4.24	27.00	47.50
ETP = -0.5	32.84	8.24	0.00	51.00
ETP = 0.0	33.80	6.75	12.00	50.00
ETP = 1.0	28.28	9.04	0.00	51.00
ETP = 3.5	29.82	13.37	0.00	50.00
Width = 0	38.86	3.79	32.50	50.00
Width = 1	30.83	11.50	0.00	51.00
Width = 2	33.05	6.47	10.00	48.50
Width = 4	26.52	9.18	0.00	51.00
Width = 6	24.05	4.70	12.00	35.00

Note: N=39

Table 3. Estimated Coefficients from Panel Regression Models.

	Model 1 (Yaniv-Foster)		Model 2		Model 3	
Constant	29.28	***	29.32	***	35.74	***
	(0.38)		(0.83)		(1.03)	
Coverage			4.03	***	4.10	***
			(0.90)		(0.86)	
|ETP|			-0.67	***	-1.56	***
			(0.26)		(0.25)	
ETP^2						
$ETP*D_{+err}$					0.74	**
					(0.37)	
Log(width)	-1.46	***	-1.19	***		
	(0.19)		(0.17)			
Width					-2.46	***
					(0.20)	
Cross-sections	39		39		39	
Observations	592		592		592	
Adj. R^2	0.21		0.33		0.44	

Statistical significance at 1% ***, 5% **, and 10% *. PCSE standard errors in parentheses.

(C) If the price of your stock turned out to be $50 in six months, which forecast you wish you had purchased when you were planning?

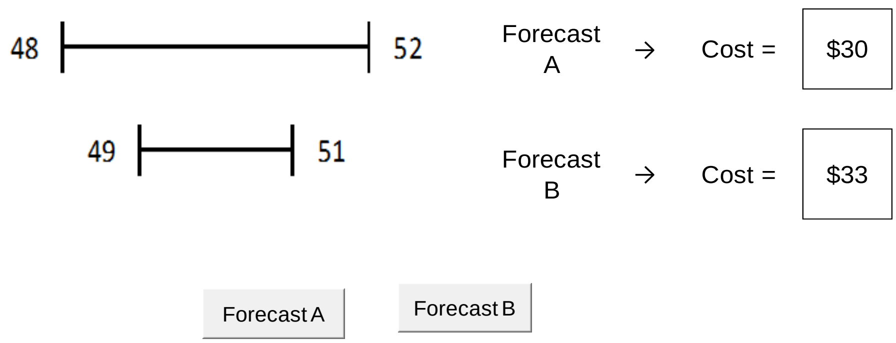

Figure 1. Sample Choice Experiment Question.