Journal of Agricultural and Applied Economics, 36,3(December 2004):605–615

Predicting Pork Supplies: An Application of Multiple Forecast Encompassing

Dwight R. Sanders and Mark R. Manfredo

Conditional efficiency or forecast encompassing is tested among alternative pork production forecasts using the method proposed by Harvey and Newbold. One-, two-, and three-quarter ahead pork production forecasts made by the United States Department of Agriculture (USDA), the University of Illinois and Purdue University Cooperative Extension Service, and those produced by a univariate time series model are evaluated. The encompassing tests provide considerably more information about forecast performance than a simple pair-wise test for equality of mean squared errors. The results suggest that at a one-quarter horizon, the Extension service forecasts encompass the competitors, but at longer horizons, a composite forecast may provide greater accuracy.

Key Words: composite forecasts, forecast encompassing, pork production

JEL Classifications: C53, Q13

Agribusiness decision-makers often have a variety of forecasts to choose from when predicting an economic variable. For instance, meat processors have meat production forecasts available from the U.S. Department of Agriculture, cooperative Extension services, and private forecasters, and they may even produce their own forecasts in-house. Therefore, it is important that agribusiness decision-makers know which forecasts contain the most information, which ones are most accurate, and which forecasts can be combined in order to provide greater accuracy. In some instances a particular forecast may be redundant, which means that all of its information is embedded in other forecasts. Knowing this allows for streamlined decision-making and can also reduce operation costs by eliminating redundant forecasting services or modeling efforts.

Traditionally, economists have relied on qualitative assessments or pair-wise quantitative methods to compare alternative forecasts (Armstrong; Hafer and Hein; Kastens, Schroeder, and Plain). For instance, Egelkraut et al. use the modified Diebold Mariano (MDM) test proposed by Harvey, Leybourne, and Newbold (1997) to compare United States Department of Agriculture (USDA) crop production forecasts to those made by two private forecasting agencies, Sparks and Leslie. For one sample period examined, they find that the mean absolute percentage errors were statistically smaller for Sparks versus Leslie and for Sparks versus the USDA. This is useful information if a forecast user is limited to choosing one, and only one, forecast. However, these results do not indicate whether the Sparks forecast contains all of the information embedded in the competing forecasts, in which case the alternative forecasts can be eliminated

---

Dwight R. Sanders is assistant professor, Department of Agribusiness Economics, Southern Illinois University, Carbondale, IL. Mark R. Manfredo is assistant professor, Morrison School of Agribusiness and Resource Management, Arizona State University, Mesa, AZ.

The authors wish to thank two anonymous reviewers of the *Journal* for their useful comments and insights.

from consideration. In this sense, pair-wise accuracy tests do little to reduce the volume of data a decision-maker must consider.

A complementary approach to pair-wise accuracy tests is to test the concept of forecast encompassing, or conditional efficiency, among alternative forecasts. A preferred forecast is said to "encompass" a competing forecast if there is not a linear combination of forecasts (i.e., a composite forecast) that would produce a smaller mean squared error than that of the preferred forecast (Mills and Pepper). Encompassing tests have been used to some extent in previous research assessing the performance of alternative livestock production forecasts. For instance, incorporating the encompassing method proposed by Harvey, Leybourne, and Newbold (1998), Sanders and Manfredo examined the performance of USDA livestock production forecasts. They found that USDA one-quarter ahead livestock production forecasts do not encompass the information contained in simple time series forecasts.[1] Therefore, their findings suggest that a decision-maker can improve forecast accuracy in this context by using a composite of the USDA and time series forecasts (Armstrong; Bates and Granger; Brandt and Bessler; Johnson and Rausser). Although Sanders and Manfredo only examine two competing forecasts (USDA and time series), Harvey and Newbold have suggested a methodology for testing "multiple forecast encompassing" for situations where three or more competing forecasts are available. As well, prior research evaluating livestock production forecasts focuses on one-quarter ahead (Sanders and Manfredo) or annual forecast horizons (Bailey and Brorsen). Often, however, decision-makers need quarterly information beyond one-step ahead. For instance, hog producers have a natural production lag of over 9 months from conception to slaughter. So for them, it is important to examine multiple-quarter ahead forecasts.

[1] In another application of the MDM and forecast encompassing techniques proposed by Harvey, Leybourne, and Newbold (1997, 1998), Wang and Bessler examine the forecasting performance of demand systems for beef, poultry, and pork both with and without the restriction of homogeneity.

Therefore, this research seeks to illustrate and apply the multiple forecast encompassing framework of Harvey and Newbold to pork production forecasts from multiple sources (USDA, expert opinion, and a time series model) over alternative forecast horizons (one-, two-, and three-quarter ahead). Thus, this research provides important information to decision-makers in the pork industry as to the weight that they should place on alternative pork production forecasts. This may allow them to reduce and simplify the information set used in the decision-making process. However, the most important contribution of this research is that it serves as a clear example of how multiple forecast encompassing could be used in other applications.

In the next section, we first present a description of the data used. Then, in the following section, we examine forecast accuracy of the pork production forecasts using traditional measures such as mean squared error. Next, a MDM test is used to determine if the differences in predictive accuracy are statistically significant. We then test for multiple forecast encompassing and also present estimates of composite forecast weights. Finally, the paper concludes with a summary and discussion of the results and suggestions for future research.

Data

This research focuses on three different sources of pork production forecasts: the USDA, the University of Illinois and Purdue University Cooperative Extension Service, and a simple time series model. The USDA releases quarterly production forecasts in their monthly World Agricultural Supply and Demand Estimates (WASDE) reports. These reports are issued between the eighth and 14th day of each month and contain a set of quarterly forecasts for commercial pork production for at least the ensuing three quarters. Three-quarter ahead forecasts have been available on a consistent basis since the first quarter of 1994. Thus, the forecasts used in this study are drawn from the first report of each quarter: January, April, July, and October. The sample covers the period from the first quarter of 1994 (1994.1) to the fourth quarter of 2002 (2002.4), resulting in 36, 35, and 34 forecasts at the one-, two-, and three-quarter horizons, respectively. This collection methodology keeps the information set available to the USDA forecasters consistent with that of the Extension service forecasters.

Purdue University, in cooperation with the University of Illinois, publishes quarterly pork production forecasts in their Livestock Price Outlook. Production forecasts for one-, two-, and three-quarter ahead forecasts are collected from the first quarter of 1994 through the fourth quarter of 2002 to match up with the USDA forecasts. Over this sample, Chris Hurt of Purdue University authored the Livestock Price Outlook. These reports are usually issued in the first 2 weeks of the quarter—at approximately the same time as the WASDE reports and shortly after the USDA's quarterly Hogs & Pigs report.

Finally, a simple time series model is used to generate out-of-sample forecasts over the same interval. Granger suggests the use of univariate time series models as a low-cost standard of comparison for forecasters. In this spirit, the natural logarithm of actual pork production is seasonally differenced and modeled in a Box-Jenkins framework. In the pre-forecast sample (1975.1 to 1993.4), an AR(4) model fit the data well and the residual autocorrelation and partial autocorrelation functions were not statistically significant out to eight lags. Therefore, we use this very simple AR(4) model to generate one-, two-, and three-quarter ahead forecasts. Specifically, the model is estimated from 1975.1 through 1993.4. This model is used to make one-, two-, and three-quarter ahead forecasts (1994.1, 1994.2, and 1994.3, respectively). Then, the model is estimated from 1975.1 through 1994.1, and this model is used to forecast 1994.2, 1994.3, and 1994.4. This process continues through 2002.4. Although the data used to estimate the model are continually updated, the model is not respecified. Therefore, the time series model here does not take into consideration learning, which likely is an influence in both the USDA and Extension forecasts. Indeed, USDA and Extension forecasters have the ability to adjust their forecasting techniques over time regardless of the method (e.g., econometric models, time series models, or pure expert opinion). Although USDA and Extension forecasters have the ability to adjust forecast methods over time, they still may not be making full use of the information provided by a simple time series model. Additionally, a time series model that is not respecified over time avoids potentially subjective efforts to find the "best" time series model, and perhaps "data snooping" in the process. It also produces an easily replicable set of results, which would not be the case if the model were continually respecified.[2]

The absolute level of meat production demonstrates strong seasonality and trends. Therefore, to assure stationarity in the variables, the analysis focuses on seasonal differences defined as the change in production from the same quarter of the prior year. Furthermore, the data are converted to log levels such that the seasonal differences represent percent changes from the same quarter of the prior year. For example, let A_t equal the level of production in quarter t, and let F_t equal the forecasted production for quarter t. The variables of interest are thus defined as the change in actual production, $AP_t = \ln(A_t/A_{t-4})$, and the forecasted change in production, $FP_t = \ln(F_t/A_{t-4})$, such that the change represents the percent change in quarterly meat production from the prior year. Organizing the data in this manner provides time series that are consistent with those used by the trade and most industry analysts (Hurt), and it is consistent with prior research (Kastens, Schroeder, and Plain). As well, previous research suggests that pork production forecasts may contain a slight bias (Sanders and Manfredo). Therefore, following Harvey and Newbold, the forecast error series are demeaned prior to implementing the tests for forecast encompassing.

[2] Bayesian methods represent an alternative approach that has been advocated to account for the ability of forecasters to adapt or learn over time (Bessler and Chamberlain; Bunn).

Table 1. Forecast Accuracy Measures

Forecast Horizon	USDA	Extension	Time Series
Panel A: RMSE			
One quarter	2.798	2.541	3.037
Two quarters	3.281	3.002	3.301
Three quarters	4.249	3.634	3.763
Panel B: MAE			
One quarter	2.034	1.926	2.426
Two quarters	2.445	2.313	2.782
Three quarters	3.227	2.664	3.119
Panel C: Theil's U			
One quarter	0.575	0.522	0.624
Two quarters	0.665	0.608	0.669
Three quarters	0.851	0.728	0.753

Note: The unit of measure for root mean squared error (RMSE) and mean absolute error (MAE) are percentages. For example, the RMSE for USDA one-quarter ahead forecasts is 2.798%.

Forecast Accuracy

Forecast accuracy is evaluated with three traditional measures: root mean squared error (RMSE), mean absolute error (MAE), and Theil's U. The results are presented in Table 1. Based on RMSE (Table 1, Panel A), the Extension forecasts have the smallest squared error at all forecast horizons. For instance, at two-quarters ahead, the Extension forecasts produce an RMSE of 3.002% followed by 3.281% and 3.301% for the USDA and time series models, respectively. Interestingly, at a three-quarter horizon, the time series model produces a smaller RMSE than the USDA. The results for the MAE (Table 1, Panel B) are comparable with those using the RMSE. That is, the Extension forecasts are the most accurate at all horizons. The USDA has a smaller MAE than the time series model at one and two steps ahead, but not at three-quarters ahead.

Theil's U normalizes forecast errors by the volatility of the underlying series. Theil's U has a lower bound of zero for perfect forecasts, and it takes a value of unity for naive "no change" forecasts (Leuthold). As expected, all of the forecasts provide performance superior to a "no change" naive alternative (Table 1, Panel C). Consistent with the previous results, the Extension forecasts provide the most improvement at all horizons. The USDA is slightly better than the time series model at two-quarters ahead, and the time series outperforms the USDA at a three-quarter horizon. Although these casual comparisons of forecast accuracy measures are informative, it is more useful to compare them in a statistical sense. Here, we use the method proposed by Diebold and Mariano and modified by Harvey, Leybourne, and Newbold (1997).

Diebold and Mariano introduced a general method for testing the equality of forecast errors. Given two time series of h-step ahead forecast errors (e_{1t}, e_{2t}) for $t = 1, \ldots, n$, and a specified loss function $g(e)$, the null hypothesis of equal expected forecast performance is $E[g(e_{1t}) - g(e_{2t})] = 0$. For h-step ahead forecasts, the MDM test is based on the sample mean $(\bar{d})$ of $d_t = g(e_{1t}) - g(e_{2t})$ with appropriate adjustments for $h - 1$ autocorrelation. In particular,

$$(1) \quad \text{MDM} = \left[\frac{n + 1 - 2h + n^{-1}h(h - 1)}{n}\right] \times \left[n^{-1}\left(\hat{\gamma}_0 + 2\sum_{k=1}^{h-1}\hat{\gamma}_k\right)\right]^{-1/2} \bar{d},$$

where

$$\hat{\gamma}_k = n^{-1}\sum_{t=k+1}^{n}(d_t - \bar{d})(d_{t-k} - \bar{d})$$

is the estimated kth autocovariance of d_t, and

Table 2. Modified Diebold Mariano (MDM) Test for Equality of Mean Squared Errors

	Extension	Time Series
Panel A: one-quarter ahead		
USDA[a]	0.237	0.512
Extension		0.231
Panel B: two-quarters ahead		
USDA	0.239	0.974
Extension		0.587
Panel C: three-quarters ahead		
USDA	0.042	0.415
Extension		0.800

[a] P-value for incorrectly rejecting the null hypothesis that the MSEs are equivalent.

$\bar{d}$ is the sample mean of d_t. The MDM statistic is compared with the critical values from a t-distribution with $n - 1$ degrees of freedom. Harvey, Leybourne, and Newbold's (1997) modifications to the original Diebold Mariano statistic make it more robust in the presence of nonnormal errors. Also, given that one-, two-, and three-step ahead forecast errors are examined, $h - 1$ autocorrelation in the forecast error series is likely introduced. Thus, with this data, the MDM test is an appropriate test for differences in forecast accuracy.

The p-values from the MDM test are presented in Table 2. For one-quarter ahead forecasts (Table 2, Panel A), the null hypothesis of equal mean squared errors cannot be rejected at conventional significance levels. The results for two-quarter ahead forecasts (Table 2, Panel B) are similar in that there is not a statistically significant difference between the mean squared errors of the different forecasts. The null hypothesis of equal mean squared errors can only be rejected between the Extension forecasts and those of the USDA at a forecast horizon of three quarters (5% level). Therefore, although the Extension forecasts seem to be more accurate, the only statistically significant difference in forecast accuracy is at the three-quarter horizon when compared with the USDA. This result is not necessarily surprising given the recent evidence by Ashley that as many as 100 observations may be necessary for a 20% reduction in mean squared errors to be statistically significant at the 5% level.

Pair-wise comparisons of forecast accuracy are a vital component in assessing forecast models. But, as shown above, the results are often inconclusive and may not provide additional information to forecast users beyond indications of point accuracy. For instance, the above results suggest that the Extension forecasts are statistically more accurate at a three-quarter horizon than those made by the USDA. But, does this necessarily mean that the USDA forecasts can be ignored? Indeed, the USDA forecast may contain valuable information that can complement the Extension forecasts and may even be useful in creating a composite forecast of pork production (Brandt and Bessler; Johnson and Rausser). As pointed out by Granger and Newbold (p. 267), if the forecast errors are sufficiently uncorrelated, then a composite forecast that utilizes both the USDA and Extension forecasts may be optimal. To advance along these lines, the following section evaluates the USDA, Extension, and time series forecasts in a multiple forecast encompassing framework, which yields valuable information as to how to combine and use the separate forecasts, if at all.

Forecast Encompassing

Harvey and Newbold state that one forecast is said to encompass another if the inferior forecast's optimal weight in a composite predictor is zero. With just two competing predictors, forecast encompassing can be tested with the regression-based model

$$(2) \qquad e_{1t} = \alpha + \lambda(e_{1t} - e_{2t}) + \varepsilon_t.$$

Here, e_{1t} is the forecast error series of the preferred forecasts, and e_{2t} is the forecast error series of the competing forecasts. A test of the null hypothesis, $\lambda = 0$, is a test that the covariance between e_{1t} and $(e_{1t} - e_{2t})$ is zero. A failure to reject the null hypothesis implies that a composite forecast cannot be constructed that would result in a smaller expected squared error than using the preferred forecast by itself. Thus, the preferred forecast "encompasses" or is "conditionally efficient" with respect to the competitor (Harvey, Leybourne, and Newbold, 1998).

To gain some intuition into this test, consider some extreme examples. First, consider the case where the preferred and competing forecasts are identical. Thus, the forecast errors are identical ($e_{1t} = e_{2t}$). In this case, the competing forecast clearly provides no marginal information to the preferred, and the optimal weight in a composite forecast is trivially zero ($\lambda = 0$). Now, consider the case where the alternative forecast produces an error, e_{2t}, that is of equal size but the opposite sign of the preferred forecast error, e_{1t}. Then, the estimated λ in Equation (2) would be 0.5 and the optimal composite predictor would be an equally weighted average of the preferred and alternative forecasts.

The Extension of forecast encompassing to a multivariate case is straightforward. Using Harvey and Newbold's notation, consider the availability of K forecasts, f_{it}, ($i = 1, 2, \ldots, K - 1, K$) for the economic variable A_t. The regression-based test that f_1 encompasses $f_2, f_3, \ldots, f_{K-1}, f_K$ is expressed as follows:

$$(3) \quad e_{1t} = \alpha + \lambda_1(e_{1t} - e_{2t}) + \lambda_2(e_{1t} - e_{3t}) + \cdots + \lambda_{K-2}(e_{1t} - e_{K-1t}) + \lambda_{K-1}(e_{1t} - e_{Kt}) + \varepsilon_t,$$

where e_{1t} is the forecast error series of the preferred forecasts, and e_{Kt} is the forecast error series of the Kth competing forecast. The null hypothesis that f_1 encompasses $f_2, f_3, \ldots, f_{K-1}, f_K$ is a joint test that $\lambda_i = 0$ for all i and can be tested with a standard F-test. However, Harvey and Newbold show that the standard F-test exhibits significant size distortions in small and moderate samples with nonnormal errors. In particular, under nonnormal errors, the standard F-test is severely oversized in small and moderate samples. Additional simulation results lead the authors to recommend a Diebold Mariano–type test, similar to the MDM test presented in Equation (1), as the preferred test for multiple forecast encompassing due to good size and power in moderately large samples.

Harvey and Newbold's MDM test for multiple forecast encompassing is essentially a test that $\lambda_i = 0$ for all i in Equation (3) based on testing that the covariance between e_{1t} and $(e_{1t} - e_{it})$ is zero for all i. That is, define $d_{it} = (e_{1t} - e_{i+1,t})e_{1t}$, and $\Delta_t = [d_{1t}\ d_{2t}\ d_{3t} \cdots d_{K-1,t}]$. Then, the null hypothesis is that the vector of covariance terms, Δ_t, equals zero. Specifically, Harvey and Newbold suggest a test based on Hotelling's generalized T^2-statistic,

$$(4) \quad MS^* = (K - 1)^{-1}(n - 1)^{-1} \times (n - K + 1)\bar{d}'\hat{V}^{-1}\bar{d}.$$

In Equation (4), $\bar{d} = [\bar{d}_1\ \bar{d}_2\ \bar{d}_3 \cdots \bar{d}_{K-1}]$, $\bar{d}_i = n^{-1} \sum d_{it}$ and $\hat{V}$ is the sample covariance matrix. Furthermore, the sample covariance matrix must be adjusted to account for the implicit $(h - 1)$ dependency in h-step ahead forecasts (Harvey, Leybourne, and Newbold, 1997). Therefore, in $\hat{V}$, the (i, j)th element is defined as

$$\hat{v}_{ij} = n^{-1}[n + 1 - 2h + n^{-1}h(h - 1)]^{-1} \times \left[\sum_{t=1}^{n} (d_{it} - \bar{d}_i)(d_{jt} - \bar{d}_j) + \sum_{m=1}^{h-1} \sum_{t=m+1}^{n} (d_{it} - \bar{d}_i)(d_{j,t-m} - \bar{d}_j) + \sum_{m=1}^{h-1} \sum_{t=m+1}^{n} (d_{i,t-m} - \bar{d}_i)(d_{jt} - \bar{d}_j)\right].$$

In finite samples, MS^* is distributed as an $F_{K-1,n-K+1}$. Thus, we test for forecast encompassing using Harvey and Newbold's MS statistic in Equation (4). The MS^* statistic serves as the most appropriate test of the null hypothesis that the preferred forecast encompasses the alternatives: $\lambda_i = 0$ for all i, in Equation (3).

Harvey and Newbold's multiple forecast encompassing test presented in Equation (4) is conducted at each forecast horizon. Additionally, we use each forecast series as the "preferred" model. This allows for a more thorough understanding of the relationship among the forecast series. As pointed out by Harvey and Newbold, failure to reject the null hypothesis does not necessarily imply that the preferred forecast is strictly dominant to the competing forecasts. Rather, the forecasts may be highly correlated or similar, in which case

Table 3. Test for Multiple Forecast Encompassing

	"Preferred Forecast"		
Forecast Horizon	USDA	Extension	Time Series
One quarter[a]	0.006	0.245	0.056
Two quarters	0.236	0.272	0.101
Three quarters	0.036	0.028	0.096

[a] *P*-values from Harvey and Newbold's test for multiple forecast encompassing.

a combination of nearly identical forecasts could not improve upon any individual forecast. A failure to reject the null hypothesis may also arise due to large sample variability or potentially low power and undersizing of the test. However, rejection of the null hypothesis in the encompassing test leads to a much stronger inference: the preferred forecast can be improved by combining it with the alternatives. Given this, it is important to carefully weigh all the evidence when interpreting the results.

Table 3 presents the *p*-values for the null hypothesis that the preferred forecast encompasses the alternatives. As shown in the first row of Table 3, for a one-quarter forecast horizon, the null hypothesis of forecast encompassing is rejected at the 5% level for the USDA and at the 10% level for the time series forecasts, respectively. That is, these forecasts can be improved by combining them with the other forecast series. Importantly, the null hypothesis cannot be rejected at the 10% level for the Extension forecasts. This suggests that at a one-quarter horizon, the Extension forecasts contain all of the information in both the USDA and time series alternatives. The implication is that decision-makers interested in one-quarter ahead forecasts of pork production can focus on the Extension forecasts exclusively, thereby simplifying the information set and overall decision-making process. In many ways, this result should not be surprising given that Extension forecasters likely use USDA forecasts, or at least USDA supply and demand data, in forming their own forecasts. Indeed, Extension forecasters often make reference to USDA forecasts in the outlook information they provide. Thus, their forecasts would encompass the information provided by USDA forecasts. As well, this result is also consistent with the smaller RMSE demonstrated by the Extension forecasts in Table 1. However, it is interesting that the MDM test was unable to reject that the mean squared forecast errors were equal in a pair-wise comparison (Table 2). This may suggest that the pair-wise MDM test lacks power (Ashley) and that the encompassing test is more useful in distinguishing among out-of-sample forecasts.

The results at a two-quarter ahead horizon are not as clear (Table 3, second row). At this horizon, the null hypothesis of forecast encompassing is rejected at the 10% level (p = .101) for the time series model's forecasts, indicating that these forecasts can be improved by combining them with the USDA or Extension forecasts. However, the null hypothesis cannot be rejected for either the USDA or Extension forecasts. Therefore, there is no evidence that these two-quarter ahead forecasts can be improved by combining them with their competitors. This suggests that for two-quarter ahead forecasts, decision-makers can utilize either the USDA or Extension forecasts alone, but the time series forecasts should be used in a composite predictor.

Three-quarter ahead forecasts are presented in the bottom row of Table 3. As one might expect, forecasting pork production 6–9 months into the future is a difficult task that is best accomplished with a combination of techniques. That is, each forecast series can be improved through combinations with the other forecasts. The null hypothesis of forecast encompassing is rejected at the 5% level for the USDA and Extension forecasts and at the 10% level for the time series forecasts. This would suggest that none of the forecasts should be used independently; rather, a combination of forecasts will result in a smaller mean squared error (Bates and Granger; Brandt and Bessler; Johnson and Rausser). In the following section, we further investigate the optimal combination of forecasts.

Combining Forecasts

There are many potential methods for combining forecasts (see Bessler and Chamberlain;

Table 4. Composite Forecast Weights from Encompassing Regressions

Forecast Horizon	USDA	Extension	Time Series
One quarter	−0.164	0.953	0.211
	(0.361)[a]	(0.372)	(0.191)
Two quarters	−0.423	0.963	0.460
	(0.414)	(0.445)	(0.185)
Three quarters	−0.325	0.282	1.043
	(0.259)	(0.263)	(0.089)

[a] Standard errors are in parentheses.

Brandt and Bessler; Bunn; Granger and Newbold, p. 269). However, the logical application in this research is to simply estimate Equation (3), where the estimated λ_i's are the weights on the competing forecasts, and $1 - \Sigma\ \lambda_i$ is the weight on the preferred forecast. However, Harvey and Newbold show that the covariance estimator from estimating Equation (3) results in poorly sized test statistics under certain distributional assumptions. Regardless, the parameter estimates are not biased and may provide valuable information in regard to weighting the competing forecasts. Therefore, Equation (3) is estimated using ordinary least squares, where the h-step ahead forecasts are allowed to follow an $h - 1$ order moving average process, $\varepsilon_t = v_t - \Sigma_{i=1}^{h-1}\ \theta_i v_{t-i}$, as suggested by Harvey and Newbold. The parameter estimates in Equation (3) are not restricted to positive values. Therefore, it is possible to get negative weights for a particular competing forecast series when that forecast's errors are highly correlated with the errors of the preferred forecast (Granger and Newbold, p. 267).

The sample forecast weights estimated with Equation (3) are shown in Table 4 with the estimated standard errors in parentheses. Again, the standard errors may not be precisely estimated; therefore, care must be exercised in using the standard errors for statistical tests. The results for one-quarter ahead forecasts are consistent with the encompassing test results. The weights on the USDA and time series models are relatively small, and the estimated weight on the Extension forecast is near one. This is added evidence that Extension forecasts may be used by themselves for one-quarter ahead forecasting.

The estimated weights (Table 4, second row) help clarify the two-quarter ahead results. The weights would suggest combining the Extension and time series forecasts. The estimated weights on both of these forecast series are relatively large. It is important to note that the negative weight on the USDA forecast stems from its high correlation (correlation = 0.95) with the Extension forecasts (Granger and Newbold, p. 267). Thus, the USDA forecast is highly correlated with the Extension forecasts, but not as accurate in a mean squared error framework (Table 1). As a result, the negative weight on the USDA forecasts serves to reduce the overall weight on two similar forecasts. For instance, the forecasts for 1999.4 are 2.18% and 2.20% for the USDA and Extension forecasts, respectively, whereas the time series forecast is −1.24%. Using the weighting scheme in the second row of Table 4, we get a composite forecast of 0.626% ($-0.423 \times 2.18 + 0.963 \times 2.20 + 0.460 \times -1.24$). As can be seen in the calculation, the two similar forecasts are getting an aggregate weight of 0.540, and the more uncorrelated time series forecast gets a weight of 0.460. This suggests that at a two-quarter horizon, a simple average of the Extension and time series forecasts may provide an accurate and easily obtainable forecast for decision-makers.

Surprisingly, the composite forecast at a three-quarter horizon is heavily weighted toward the time series model. The weights for the USDA, Extension, and time series models are −0.325, 0.282, and 1.043, respectively. Again, the negative weight on the USDA forecast stems from its similarity to the Extension forecast (correlation = 0.96). The result is an aggregate negative weight on these two similar forecasts of −0.043 (0.282 − 0.325) and a weight near one, 1.043, on the time series model. This would tentatively indicate that at a forecast horizon of three quarters, the time series model might be used by itself. This is consistent with the RMSE reported in Panel A of Table 1, where the time series model generates an RMSE smaller than the USDA and nearly as small as the Extension forecasts. It can also be reconciled with the forecast encompassing results. That is, the null of forecast encompassing was rejected at the 5% level for the USDA (p = .036) and Extension (p = .028) forecasts, most likely due to an exclusion of the time series properties in the forecast series. The time series forecast, however, rejected forecast encompassing at the 10% level, but not at the 5% level.

Collectively the empirical results suggest that at one-quarter horizons, the USDA and time series forecasts can be improved upon by using a linear combination of forecasts, whereas the Extension forecasts cannot. Furthermore, the estimated composite forecast weights suggest that at this horizon, the Extension forecast may be effectively used by itself as a predictor of pork production. At a two-quarter horizon, the null hypothesis of forecast encompassing is rejected for the time series forecasts, suggesting that it may be improved through a combination with the other forecasts. Estimation of the composite forecast weights indicates that an equal weighting of Extension and time series forecasts represent a plausible alternative at this horizon. Finally, at a three-quarter horizon, neither the Extension nor USDA forecasts encompass the information contained in the competing predictions. The time series model rejects the null hypothesis at the 10% level but not at the 5% level. The estimated composite forecast weights would suggest that at this longer horizon, the time series model may be used alone as a forecast of pork supply.

Summary and Conclusions

Decision-makers often must consider multiple forecasts for an economic variable. In particular, industry participants in the food sector have pork production forecasts available from numerous sources (e.g., USDA, Extension services, and internally produced forecasts). If the decision-maker can effectively eliminate redundant forecasts, then this can narrow the information set, thereby simplifying the decision-making process and possibly reducing the cost of producing internal forecasts. The paper presents a methodology for comparing forecasts, choosing those that encompass the information in competing forecasts, and creating a combined forecast when appropriate.

Pork production forecasts provided by the USDA, Illinois and Purdue Cooperative Extension Service, and those made by a univariate time series model are compared at forecast horizons of one-, two-, and three-quarters ahead. Traditional accuracy measures suggest that the Extension service forecasts are most accurate at all horizons, usually followed by the USDA and then the time series forecasts. The exception is at a three-quarter horizon, where the time series model is more accurate than the USDA. The equality of mean squared prediction errors can only be rejected between the USDA and Extension forecasts at the three-quarter horizon. Although informative, this set of tests provides little insight as to the relative value of each forecast series.

The forecasts are further evaluated using the forecast encompassing test proposed by Harvey and Newbold. For one-quarter ahead forecasts, the encompassing test suggests that the USDA and time series model's forecasts do not encompass the information contained in the competing forecasts. However, the Extension forecasts cannot be improved by combining them with the USDA or time series forecasts. An examination of the variance minimizing weights suggests that at one-quarter horizons, the Extension forecasts may be used by themselves. For the two-quarter ahead horizon, the time series model rejects the null hypothesis of forecast encompassing, and the USDA and Extension forecasts fail to reject the null, indicating that they cannot be improved through a combined forecast. Again, the estimated composite forecast weights are evaluated and suggest that a composite forecast using Extension and the time series model is a potentially useful alternative. Finally, for three-quarter ahead forecasts, there is evidence that the time series forecasts encompass those made by the USDA and Extension economists. That is, the null hypothesis is rejected at the 5% level for the USDA and Extension forecasts, but only at the 10% level for the time series forecasts. Furthermore, the minimum variance composite forecast is heavily weighted toward the time series model. Collectively this suggests that the time series model may be used by itself as a forecast of pork supply at this horizon.

Overall, the empirical results provide the following practical recommendations for agribusiness decision-makers. For one-quarter ahead pork production forecasts, those provided by the Illinois and Purdue Extension Service are difficult to improve upon (using USDA or univariate time series models). Furthermore, a practical two-quarter ahead forecast would be a simple average of the Extension and time series forecasts. Finally, at a three-quarter horizon, the time series model's forecasts by themselves are a viable alternative. It is important to note, however, that these recommendations are limited by the forecasts examined. That is, the results and recommendations will most certainly differ if forecasts generated by other models or entities are utilized. Nonetheless, this research sets forth one possible methodology for comparing forecasts, especially when the goal is to eliminate redundant information without reducing forecast accuracy, or merely to gain further insight into the information content of alternative forecasts. Armed with this information, agribusiness decision-makers will likely save time and potentially lower the costs of their forecasting efforts.

USDA and Extension forecasters can also use these results to evaluate and potentially reexamine their forecasting methods. First, the USDA should consider supplementing or improving their current forecasting techniques. The results indicate that the USDA's forecasts can likely be improved by adding a time series component to the forecasts at all horizons. Second, Extension forecasters are doing an admirable job, especially at short horizons. The Extension forecasts are the most accurate of the three examined, and they clearly encompass the competing forecasts at a one-quarter horizon. Indeed, Extension forecasters are efficiently incorporating the information available to them (e.g., USDA forecasts and time-series properties) in forming their forecasts at the one-quarter horizon. This is an interesting phenomenon given the arguments of Meehl who suggests that experts have often been shown to provide inferior forecasts relative to those produced by statistical models. However, there is some evidence that at two- and three-quarter ahead horizons the Extension forecasts are missing some of the time series properties of the data and they may be improved with a composite forecast. The conclusion that composite forecasts may be better than any individual forecast series at longer horizons is not meant to be a critical assessment of the public worth of the outlook services provided by either the USDA or by Cooperative Extension, but it does provide additional evidence with regard to the usefulness of composite forecasting. At longer horizons, it inherently gets more difficult for experts to forecast, and simple time series models may provide a means of conveniently summarizing past data on pork supply. Indeed, given the difficulty of forecasting at long horizons, simple time series models may provide an objective alternative (Meehl). As pointed out by Clements, the reason for finding that combinations of methods fare better than individual forecasts can include model misspecification, but it may also arise from nonstationarities and locational shifts in the data that are captured better by different modeling techniques. Future research may include the use of Bayesian methods of combining forecasts (Bessler and Chamberlain; Bunn), which considers the ability of the forecaster to adapt and incorporate their prior beliefs of forecast performance in developing the weights used in a composite forecasting framework. Regardless, multiple forecast encompassing is a powerful tool for both forecasters and decision-makers to assess the relative value of alternative forecasts.

[Received August 2003; Accepted November 2003.]

References

Armstrong, J.S. *Principles of Forecasting: A Handbook for Researchers and Practitioners.* Boston: Kluwer Academic Publishers, 2001.

Ashley, R. "Statistically Significant Forecasting Improvements: How Much Out-of-Sample Data

Is Likely Necessary?" *International Journal of Forecasting* 19(2003):229–39.

Bailey, D.V., and B.W. Brorsen. "Trends in the Accuracy of USDA Production Forecasts for Beef and Pork." *Journal of Agricultural and Resource Economics* 23(1998):515–25.

Bates, J., and G.W.J. Granger. "The Combination of Forecasts." *Operational Research Quarterly* 20(1969):451–68.

Bessler, D.A., and P.J. Chamberlain. "Composite Forecasting with Dirichlet Priors." *Decision Sciences* 19(1988):771–81.

Brandt, J.A., and D.A. Bessler. "Composite Forecasting: An Application with U.S. Hog Prices." *American Journal of Agricultural Economics* 63(1981):135–40.

Bunn, D.W. *The Synthesis of Forecasting Models in Decision Analysis.* Basel, Switzerland: Birkhauser, 1978.

Clements, M.P. "Some Possible Directions for Future Research." *International Journal of Forecasting* 19(2003):1–3.

Diebold, F.X., and R.S. Mariano. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13(1995):253–63.

Egelkraut, T.M., P. Garcia, S.H. Irwin, and D.L. Good. "An Evaluation of Crop Forecast Accuracy for Corn and Soybeans: USDA and Private Information Agencies." *Journal of Agricultural and Applied Economics* 35(2003):79–95.

Granger, C.W.J. "Can We Improve the Perceived Quality of Economic Forecasts?" *Journal of Applied Econometrics* 11(1996):455–73.

Granger, C.W.J., and P. Newbold. *Forecasting Economic Time Series,* 2nd ed. New York: Academic Press, 1986.

Hafer, R.W., and S.E. Hein. "Comparing Futures and Survey Forecasts of Near-Term Treasury Bill Rates." *Federal Reserve Bank of St. Louis Review* 71(May/June 1989):33–42.

Harvey, D., S. Leybourne, and P. Newbold. "Testing the Equality of Prediction Mean Squared Errors." *International Journal of Forecasting* 13(1997):281–91.

———. "Tests for Forecast Encompassing." *Journal of Business and Economic Statistics* 16(1998):254–59.

Harvey, D.I., and P. Newbold. "Tests for Multiple Forecast Encompassing." *Journal of Applied Econometrics* 15(2000):471–82.

Hurt, C. "Hog Producers Gearing Up For Expansion." *Livestock Price Outlook.* Urbana, IL: Purdue University and University of Illinois, October 2000.

Johnson, S.R., and G. Rausser. "Composite Forecasting in Commodity Systems." *New Directions in Econometric Modeling and Forecasting in U.S. Agriculture.* G. Rausser, ed. New York: North Holland Publishing, 1982.

Kastens, T.L., T.C. Schroeder, and R. Plain. "Evaluation of Extension and USDA Price and Production Forecasts." *Journal of Agricultural and Resource Economics* 23(1998):244–61.

Leuthold, R.M. "On the Use of Theil's Inequality Coefficients." *American Journal of Agricultural Economics* 57(1975):344–46.

Meehl, P.E. *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence.* Minneapolis: University of Minnesota Press, 1954.

Mills, T.C., and G.T. Pepper. "Assessing the Forecasters: An Analysis of the Forecasting Records of the Treasury, the London Business School and the National Institute." *International Journal of Forecasting* 15(1999):247–57.

Sanders, D.R., and M.R. Manfredo. "USDA Production Forecasts for Pork, Beef, and Broilers: An Evaluation." *Journal of Agricultural and Resource Economics* 27(2002):114–27.

U.S. Department of Agriculture. *World Agricultural Supply and Demand Estimates.* Internet site: http://www.usda.gov/oce/waob/wasde/wasde.htm (Accessed July 15, 2003).

Wang, Z., and D.A. Bessler. "The Homogeneity Restriction and Forecasting Performance of VAR-type Demand Systems." *Journal of Forecasting* 21(2002):193–206.