

Forecast Encompassing and Futures Market Efficiency: The Case of Milk Futures

Mark R. Manfredi
Assistant Professor
Morrison School of Agribusiness and Resource Management
Arizona State University
Email: manfredi@asu.edu

Dwight R. Sanders
Assistant Professor
Department of Agribusiness Economics
Southern Illinois University
Email: [dwights@siu.edu](mailto:dwrights@siu.edu)

*Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting,
Denver, Colorado, August 1-4, 2004*

Copyright 2004 by Mark R. Manfredi and Dwight R. Sanders. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Forecast Encompassing and Futures Market Efficiency: The Case of Milk Futures

Abstract

The traditional necessary condition for futures market inefficiency is the existence of alternative forecasting methods that produce mean squared forecast errors smaller than the futures market. Here, a more exacting requirement for futures market efficiency is proposed—forecast encompassing. Using the multiple forecast encompassing procedure of Harvey and Newbold, forecast encompassing is tested with the CME fluid milk futures contract. Time series models and experts at the USDA provide the competing forecasts. The results suggest that the CME fluid milk futures do not encompass the information contained in the USDA forecasts at a two-quarter forecast horizon. While the competing forecasts generate positive revenues, it is unlikely that trading returns would exceed transaction costs in this relatively new futures market.

Keywords: forecast encompassing, market efficiency, milk futures

Introduction

Producers and users of agricultural commodities often incorporate futures prices in economic decision making (Gardner; Hurt and Garcia). Indeed, if futures prices are used as commodity price forecasts, it is critical to understand their performance since economic welfare is enhanced by optimal and efficient forecasts (Stein; Kenyon, Jones, and McGuirk). Given the importance of this issue, a large and rich literature has developed over recent decades evaluating the forecasting accuracy of futures markets (Tomek).

Futures forecasting efficiency has been examined for numerous markets using a variety of forecast procedures. It is commonly stated that a necessary, but not sufficient condition to reject futures market efficiency is that competing forecast models produce smaller mean squared forecast errors than futures-based forecasts (Leuthold, Junkus, and Cordier, p. 116).¹ The typical assertion is that if the futures market provides the smallest mean squared forecast error, then the necessary condition for pricing inefficiency is not met; therefore, one cannot use the competing forecast model to generate trading profits. This necessary condition for futures market efficiency has been tested in the grain (Rausser and Carter), livestock (Garcia, Leuthold, Fortenbery, and Sarassoro; Martin and Garcia), energy (Ma), and financial (Leitch and Tanner; Hafer and Hein) futures markets. In this context, futures forecasts have been compared to those produced by time series and econometric models (Leuthold, Garcia, Adam, and Park), forecasts generated by commercial services (Just and Rausser), and other market experts (Irwin, Gerlow, and Liu). The overall results of these studies are mixed depending on the markets examined and alternative forecasting methods (Garcia, Hudson, and Waller). Generally speaking, futures pricing efficiency has been rejected most often using *ex post* forecasts generated by the

researchers' own models and in the livestock markets (Irwin, Gerlow, and Liu). For example, Irwin, Gerlow, and Liu indicate that there is evidence of forecast inefficiency in the livestock markets, especially at longer forecast horizons, when futures forecasts are compared to out-of-sample forecasts generated *ex post* by econometric or time series methods (Leuthold and Hartman; Leuthold, Garcia, Adam, and Park).² In contrast, studies that examine *ex ante* forecasts produced by experts in real-time generally do not reject forecast efficiency (Bessler and Brandt; Irwin, Gerlow, and Liu). In either case, the statistical criteria for forecast efficiency rests on the futures market producing a mean squared error smaller than those of competing forecasts (Leuthold, Garcia, Adam, and Park)

However, as stated by Harvey, Leybourne, and Newbold (1998), finding that forecasts (e.g., futures forecasts) are significantly better than those of a competitor should not “induce complacency” (p. 254). It is entirely possible that a forecast can have a mean squared error smaller than a competitor, but if that forecast does not “encompass” all the information in the competing forecast, then it is not conditionally efficient. In this light, the traditional necessary condition of having the smallest mean squared error is not stringent enough when testing for futures market efficiency. A higher hurdle—forecast encompassing—should be cleared in order to make any definitive arguments concerning futures market efficiency.

Given the arguments of Harvey, Leybourne, and Newbold (1998), the overall objective of this research is to illustrate that the accepted mean squared error necessary condition is not stringent enough and may lead to low power against the null hypothesis of forecast efficiency. We suggest that an efficient futures market must do more than produce the smallest mean squared forecast error. Instead, a futures forecast must meet a more exacting criterion—it must encompass all competing forecasts. Thus, this research introduces forecast encompassing as a

more exacting necessary condition for futures market efficiency. In doing this, a direct application of the encompassing principle is provided using *ex ante* forecasts produced by market experts as well as out-of-sample forecasts produced by univariate time series models over alternative forecast horizons.

This research improves on previous futures market efficiency studies in a number of ways. First, the traditional necessary condition for futures market efficiency is refined and extended. That is, this research proposes the preferred necessary condition that a futures market forecast must encompass the information contained in all competing forecasts. Second, this refined criteria for futures market efficiency is tested in the multiple forecast encompassing framework of Harvey and Newbold, which facilitates the simultaneous testing of a futures forecasts against numerous competitors. In essence, this research brings together two somewhat disparate yet related lines of literature: forecast encompassing and futures market efficiency. Third, the empirical tests of forecast efficiency used in this research are direct tests, not indirect tests such as those used in prior studies. For example, prior research such as Irwin, Gerlow, and Liu, uses futures prices to formulate a forecast for a quarterly average cash price, which is then compared to quarterly forecasts produced by an expert. This methodology is indirect because it converts a futures forecast to a quarterly average price. However, the futures market is not attempting to forecast a quarterly average price. Rather, it is forecasting a very specific delivery-time and delivery-quality price. Thus, the focus should be on the delivery-time cash price of the underlying commodity, not a quarterly average price of perhaps a non-par delivery product. Therefore, to make the empirical tests direct, this research maps expert opinion forecasts specifically to the delivery-time price underlying the futures contract (not visa-versa). Finally, this research examines the efficiency of a relatively new futures market (fluid milk), as well as a

breadth of forecast sources (USDA experts and time series), and forecast horizons (up to three quarters ahead) that have not been previously examined in a single study.

The remainder of the paper is structured as follows. First, an illustration of the mean squared error necessary condition is provided, focusing on how this measure can be misleading in testing for futures market efficiency. Second, the concept of pricing efficiency is presented in the context of forecast encompassing. Specifically, it is shown that the theoretical conditions for market efficiency are consistent with the traditional test for forecast encompassing. Next, the empirical methodology for testing futures market efficiency in a multiple forecast encompassing framework is presented and applied to milk futures, with the results compared to that of accuracy tests based on mean squared error criteria. These results are further examined vis-à-vis the sufficient condition for futures market efficiency that risk-adjusted trading profits cannot be garnered through the use of a systematic trading rule. Finally, conclusions are drawn regarding the need to strengthen the necessary condition for futures market efficiency to include forecast encompassing.

Mean Squared Error and Forecast Encompassing—Testing for Market Efficiency

Problems with MSE

Mean squared error (MSE) is used extensively to evaluate the forecasting performance of futures markets. Early studies relied on casual comparisons of MSE (Leuthold) while more recent studies have examined the statistical difference in forecast errors (Irwin, Gerlow, and Liu). In these studies, MSE is defined as:

$$\text{MSE} = \frac{1}{N} \sum_{t=1}^N (f_t - f_{t-n})^2$$

where, f_{t-n} is the futures price forecast n periods prior to time period t , f_t is the corresponding realized futures price, and N is the number of out-of-sample observations of both the forecasted and realized futures price. As stated previously, the standard necessary condition for futures market efficiency is that no competing forecast (e.g., a time series, econometric, or expert opinion forecast) provides a smaller MSE than the futures market forecast. However, differences in MSE among competing forecasts are often subtle, thus leading the researcher to wonder if differences in MSE are due only to chance. Although significant advances have been made in evaluating the statistical difference in prediction errors (Diebold and Mariano; Harvey, Leybourne, and Newbold, 1997), stating the necessary condition for futures market inefficiency strictly in a comparative MSE framework is potentially misleading. The following intuitive example illustrates how the MSE necessary condition is flawed.

Consider the following simple counter example of the MSE necessary condition. Let the true data generating process be represented by: $f_t = 0.9x_{1,t-n} + 0.1x_{2,t-n} + \varepsilon_t$, where f_t is the realized price, $x_{1,t-n}$ and $x_{2,t-n}$ are elements of the information set available at time $t-n$, ε_t is an i.i.d. white noise error, the $\text{var}(x_1) = \text{var}(x_2) = \text{var}(x)$, and $\text{cov}(x_1, x_2) = 0$. Now, assume that the futures market generates forecasts as $f_{t-n} = 0.9x_{1,t-n}$ and an alternative model's forecast is $f_{t-n}^A = 0.1x_{2,t-n}$. Then, the futures forecast error is $e^f = f_t - f_{t-n} = 0.1x_{2,t-n} + \varepsilon_t$ with $\text{var}(e^f) = 0.01\text{var}(x) + \text{var}(\varepsilon)$ and the alternative forecast error is $e^a = f_t - f_{t-n}^A = 0.9x_{1,t-n} + \varepsilon_t$ with $\text{var}(e^a) = 0.81\text{var}(x) + \text{var}(\varepsilon)$. It is clear in this example that the futures market will generate the smallest forecast error variance. But, it is equally clear that the futures market is not technically efficient. That is, it does not encompass the information contained in the alternative forecast. Therefore, a trader armed with the alternative model could conceivably use it to extract trading profits from the futures market. Given this counter example, the traditional MSE necessary condition for futures market

efficiency is incomplete, and forecast encompassing is proposed as a more exacting necessary condition. An examination of the theoretical constructs of market efficiency illustrates that the conditions for futures market efficiency are indeed more congruent with tests of forecast encompassing.

Pricing Efficiency

Pricing efficiency in the futures market proposes that if information is used efficiently and there is no risk premium, then the return for holding a futures contract from time $t-n$ to time t ($f_t - f_{t-n}$) is uncorrelated with the available information set at $t-n$ (Ω_{t-n}). Or, as pointed out by Kaminsky and Kumar, the futures forecasting error must be orthogonal to variables in the information set. This condition can be tested in the following regression equation:

$$f_t - f_{t-n} = \alpha + \beta X_{t-n} + \varepsilon_t \quad , \quad (1)$$

where, f_t is the expiration-time futures (cash) price, f_{t-n} is the futures price (forecast) at n periods prior to expiration, and X_{t-n} is a vector of variables, x_{t-n} , each an element of the available information set Ω_{t-n} .

Certainly the information set Ω_{t-n} includes other public forecasts of f_t . For example, assume that X_{t-n} consists of an alternative forecast f_{t-n}^A . Then, the null hypothesis of efficiency would imply that $\beta=0$ in the following model:

$$f_t - f_{t-n} = \alpha + \beta(f_{t-n} - f_{t-n}^A) + \varepsilon_t \quad . \quad (2)$$

That is, the change in the futures price, $f_t - f_{t-n}$, should not be related to the difference between the futures-based forecast, f_{t-n} , and the alternative forecast, f_{t-n}^A . Adding and subtracting βf_t to the right hand side of Equation (2) and simplifying yields:

$$f_t - f_{t-n} = \alpha + \beta[(f_t - f_{t-n}) - (f_t - f_{t-n}^A)] + \varepsilon_t, \text{ or} \quad (3a)$$

$$e_t = \alpha + \beta[(e_t - e_t^A)] + \varepsilon_t, \quad (3b)$$

where, e_t is the futures' forecast error, $f_t - f_{t-n}$, and e_t^A is the alternative forecast's error, $f_t - f_{t-n}^A$. Importantly, Equation (3b) is equivalent to Harvey, Leybourne, and Newbold's (1998) test for forecast encompassing. So, the traditional method of testing for smaller mean squared errors in forecast efficiency studies may lead to the wrong conclusions. It is the incremental information contained in competing forecasts, as tested in an encompassing framework, that is most consistent with the traditional definition of futures market efficiency.

Forecast Encompassing

Specifically, Harvey and Newbold state that one forecast encompasses 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 following regression-based model:

$$e_{1t} = \alpha + \lambda(e_{1t} - e_{2t}) + \varepsilon_t. \quad (4)$$

where, e_{1t} is the forecast error series of the preferred forecasts, and e_{2t} is the forecast error series of the competing forecasts.³ 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 a composite forecast cannot be constructed from the two series that would result in a smaller expected squared error than using the preferred forecast by itself. In this case, the preferred forecast is said to “encompass” the competing forecast. That is, the preferred forecast is “conditionally efficient” with respect to the competitor (Harvey, Leybourne, and Newbold, 1998).

A closer inspection of Equation (4) reveals the strength of the forecast encompassing test as applied to market efficiency. As indicated by Granger and Newbold (p. 267), in the context of Equation (4),

$$\lambda = \frac{\sigma_{e_1}^2 - \rho_{e_1e_2} \sigma_{e_1} \sigma_{e_2}}{\sigma_{e_1}^2 + \sigma_{e_2}^2 - 2\rho_{e_1e_2} \sigma_{e_1} \sigma_{e_2}} \quad \text{and} \quad \lambda \geq 0 \text{ iff } \frac{\sigma_{e_1}}{\sigma_{e_2}} \geq \rho_{e_1e_2},$$

where, σ^2 , σ , and ρ are the variance, standard deviation and correlation, respectively, among error series from the preferred, e_1 , and competing, e_2 , models. Importantly, the traditional MSE necessary condition is only concerned with the preferred (futures) forecast providing a forecast error variance smaller than the competing forecast ($\sigma_{e_1} < \sigma_{e_2}$). Whereas, the forecast encompassing test implicitly acknowledges that a seemingly inferior competing forecast ($\sigma_{e_1} < \sigma_{e_2}$) has incremental information if the portion of the forecasted variable left unexplained by the preferred forecast, e_1 , is sufficiently uncorrelated to the part left unexplained by the competing forecast, e_2 , ($\rho_{e_1e_2} < 1$). In this case, the preferred forecast does not encompass all of the information contained in the competing forecast, and λ is > 0 in Equation (4) because of the diversification benefits provided by the competing forecast. This reasoning can also be extended to the case of multiple forecast encompassing.

The extension of the forecast encompassing framework presented in equation (4) to a multivariate case is straightforward. Using notation presented in Harvey and Newbold, consider

the existence of K forecasts, f_{it} , ($i= 1,2,\dots,K-1,K$) for the economic variable A_t . The regression-based test that f_1 encompasses $f_2, f_3, \dots, f_{K-1}, f_K$ is expressed as follows:

$$e_{1t} = \alpha + \lambda_1(e_{1t} - e_{2t}) + \lambda_2(e_{1t} - e_{3t}) + \dots + \lambda_{K-2}(e_{1t} - e_{K-1t}) + \lambda_{K-1}(e_{1t} - e_{Kt}) + \varepsilon_t, \quad (5)$$

where e_{1t} is the forecast error series of the preferred forecast, and e_{Kt} is the forecast error series of the K th competing forecast. The null hypothesis that f_1 encompasses $f_2, f_3, \dots, f_{K-1}, f_K$ 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 non-normal errors. This is particularly true in the presence of non-normal errors caused by dependence in the forecast error series. Additional simulation results led Harvey and Newbold to develop and urge the use of Diebold Mariano-type tests for forecast encompassing.

Diebold-Mariano Tests

Harvey, Leybourne, and Newbold (1997) originally proposed a modification of the Diebold-Mariano test for differences in MSE to account for non-normal distributions of the forecast error series. Specifically, the modified Diebold-Mariano test (MDM) considers two time series of h -step ahead forecast errors (e_{1t}, e_{2t}) , for $t = 1, \dots, n$, and a specified loss function $g(e)$, with the null hypothesis of equal expected forecast performance being $E[g(e_{1t}) - g(e_{2t})] = 0$. 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,

$$\text{MDM} = \left[\frac{n+1-2h+n^{-1}h(h-1)}{n} \right] \left[n^{-1}(\hat{\gamma}_0 + 2\sum_{k=1}^{h-1} \hat{\gamma}_k) \right]^{-\frac{1}{2}} \bar{d}, \quad (6)$$

where $\hat{\gamma}_k = n^{-1} \sum_{t=k+1}^n (d_t - \bar{d})(d_{t-k} - \bar{d})$ is the estimated k th autocovariance of d_t , and \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 (1997) recommend the MDM test for testing differences in forecast accuracy measures; thus, it is used in this research.

Harvey, Leybourne, and Newbold (1998) also propose an extension to the MDM test for use in forecast encompassing. Specifically, they extend the MDM test to examine pairwise tests of forecast encompassing by defining $d_t = e_{1t} (e_{1t} - e_{2t})$ and \bar{d} as the sample mean of d_t , where e_{1t} and e_{2t} are defined as in equation (4). Here, the MDM is simply testing for a zero covariance between e_{1t} and $(e_{1t} - e_{2t})$, or that $\lambda = 0$ in Equation (4). The natural extension set forth by Harvey and Newbold for multiple forecast encompassing is a MDM-type test where $\lambda_i = 0$ for all i , in equation (5), which is based on testing that the covariance between e_{1t} and $(e_{1t} - e_{it})$ is zero for all i . In developing this test, they define the difference between the forecast errors of the preferred and competing forecasts as $d_{it} = (e_{1t} - e_{i+1,t})e_{1t}$, with $\Delta_t = [d_{1t} d_{2t} d_{3t} \dots d_{K-1,t}]$. Thus, the null hypothesis is that the vector of covariance terms, Δ_t , equals zero. Given this, Harvey and Newbold suggest a test statistic (MS^*) based on Hotelling's generalized T^2 -statistic:

$$MS^* = (K-1)^{-1}(n-1)^{-1}(n-K+1) \bar{d}' \hat{V}^{-1} \bar{d}, \quad (7)$$

where, $\bar{d} = [\bar{d}_1 \bar{d}_2 \bar{d}_3 \dots \bar{d}_{K-1}]$, $\bar{d}_i = n^{-1} \sum d_{it}$, 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). So, 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 $F_{K-1, n-K+1}$.

According to Harvey and Newbold, 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 (5) due to good size and power in moderately large samples. 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 in which case a combination of nearly identical forecasts could not produce a smaller mean squared error relative to an individual forecast. As well, failure to reject the null may arise due to large sample variability or potentially low power and under-sizing of the test. However, rejection of the null in the encompassing test leads to a much stronger inference. In the case of rejection, it suggests that the preferred forecast does not contain the marginal information of the competing forecasts. In the following section, forecast encompassing is tested using Harvey and Newbold's MS^* statistic with the futures market forecast designated as the "preferred" model, and time series and expert opinion forecasts designated as "competing" models.

Empirical Methodology

Given the importance and interest in the pricing efficiency of futures markets as a topic of inquiry, numerous studies have examined the efficiency of agricultural futures markets. Indeed, nearly every agricultural futures contract that is listed by an exchange today has been examined in some context (Garcia, Hudson, and Waller). However, to date, there has not been an examination of the efficiency of the relatively new fluid milk futures contracts traded at the Chicago Mercantile Exchange (CME). Initially launched as a deliverable contract in January of 1996, the CME fluid milk futures switched to cash settlement with the May 1997 contract. As well, the volume of Class III milk futures traded has greatly increased since their launch, with average daily volume increasing 85% in 2003 to 759 contracts.

In addition to being relatively new, the CME's Class III milk futures contract possesses properties that make it desirable for testing futures market efficiency. First, these contracts are cash settled to an announced price by the USDA. Thus, from a forecasting perspective, the milk futures price reflects the market's expectation for the average Class III milk price for the month. Specifically, these cash settled futures have a monthly expiration calendar, where the contract cash settles to the USDA announced average Class III price (formerly the Basic Formula Price) for the month. For instance, the July 2003 contract's last trading day was July 31, when it closed at \$11.75 per hundredweight. The contract cash settled the following day, August 1st, to the USDA announced Class III average milk price for July of \$11.78 per hundredweight. This cash settlement process is very conducive for testing the necessary conditions for futures market efficiency. For example, as pointed out by Hranaiova and Tomek, with delivery-settled futures contracts the embedded delivery options can create uncertainty as to what deliverable subset the market is pricing. Thus, changes in the cheapest to deliver quality, grade, and location could

confound the results of forecast efficiency studies that use delivery-settled futures. With cash settled fluid milk futures, there is no doubt that the futures market is forecasting the value of the USDA's announced average Class III milk price for the contract month. This provides a precisely defined set of forecasts and realized values to examine the necessary condition of futures market efficiency that previous studies lack.

In examining the necessary conditions for futures market efficiency, three sets of forecasts are used in predicting the USDA's announced Class III price: futures forecasts, forecasts generated from simple time series models, and expert opinion forecasts. In understanding the data setup and forecast horizons established, it is best to first describe the expert opinion forecasts. The USDA releases quarterly forecasts for the "all milk" price in their monthly *World Agricultural Supply and Demand Estimates (WASDE)* reports. These reports are issued between the 8th and 14th of each month and contain a set of quarterly price forecasts for the ensuing three quarters. The forecasts used are drawn from the first report of each quarter (January, March, July, and October) from the third quarter of 1997 (1997.3) through the second quarter of 2003 (2003.2), resulting in 24 *ex ante* forecast periods. Since the USDA releases forecasts for up to three quarters ahead, the sample consists of 24 one-step ahead, 23 two-step ahead, and 22 three-step ahead forecasts. An issue with this data set is appropriately mapping the USDA "all milk" price forecast to the quarterly average Class III price. This is accomplished by estimating a simple log-linear relationship between the USDA's quarterly average Class III price and the all milk price. Specifically, the natural logarithm of the Class III price is regressed against the natural logarithm of the all milk price, quarterly dummy variables, and a time trend. This provides a mapping function from quarterly all milk prices to Class III prices. This relationship is estimated with historical data up to the forecast date, and then the *WASDE*

forecasts are substituted into the relationship to get a Class III forecast for each horizon.

Importantly, the mapping relationship is estimated using data only up to the beginning of each quarterly forecast interval—preserving the *ex ante* nature of the forecasts.

Consistent with the release of the USDA expert forecasts, the futures-based forecast is calculated by averaging the three contract months that comprise the appropriate calendar quarter. Since the futures price reflects a forecast for a monthly average price (i.e., the announced Class III price), averaging these three contracts together results in a forecast for the quarterly average price that is consistent with the USDA expert opinion forecasts. Specifically, the futures-based forecast is compiled from settlement quotes taken from the day prior to the morning release of the *WASDE* report. This process yields 24 one-step ahead, 23 two-step ahead, and 22 three-step ahead *ex ante* forecasts over the span of 1997.3 to 2003.2.

Finally, two time series models are used to generate out-of-sample forecasts over the sample interval (1997.3 through 2003.2). Granger suggests the use of univariate time series models as a low-cost standard of comparison for forecasters. In the first time series model (TS-1), the natural logarithm of the USDA quarterly Class III milk price index is seasonally differenced, $\ln(p_t/p_{t-12})$, and modeled in an Box-Jenkins framework. In the pre-forecast sample (1990.1 to 1997.2) an ARMA(4,4) model fit the data well and the residual autocorrelation and partial autocorrelation functions were not statistically significant out to eight lags.⁴ Osborn, Heravi, and Birchenhall find that for highly seasonal data, annual differenced models may provide more accurate forecasts at long horizons. But, conventional first differences with the inclusion of monthly dummy variables may be more accurate at short horizons. Therefore, a second time series model (TS-2) is fit to the first differences of the natural log of the quarterly average price, $\ln(p_t/p_{t-1})$. The monthly log-relative price changes are regressed against a set of

quarterly dummy variables and the residuals are then modeled using standard Box-Jenkins techniques. The final specification includes the monthly dummy variables and an MA(4) process on the error terms.

The time series models are used to generate forecasts from one- to three-quarters ahead. Specifically, the models are estimated from 1990.1 through 1997.2. This model is used to make forecasts out for three quarters. Then, the model is estimated from 1990.2 through 1997.3, and is used to forecast from 1997.4 through 1998.2. While the most recent thirty quarters of data are used to estimate and update the model, the models are not re-specified. As pointed out by Irwin, Gerlow, and Liu, tests of forecast efficiency that rely on *ex post* model generated forecasts have tended to reject forecast efficiency. This could be due to *ex post* fitting of the data or using techniques that were not available to the market over the forecast interval (Timmerman and Granger). Here, easily replicable and simple time series specifications are employed. These procedures were widely available in standard econometric packages over the sample period.

The forecasts (futures based, expert opinion, and time series) are first evaluated using the traditional forecast accuracy measure of root mean squared error. Under the traditional necessary condition for market inefficiency, if either of the time series or expert opinion forecasts produce more accurate forecasts than the futures forecasts, the CME's Class III milk futures would be considered potentially inefficient. In addition to casual comparisons of mean squared error, the MDM procedure tests for statistical differences in forecast accuracy (Harvey, Leybourne, and Newbold, 1997). The more stringent test of pricing efficiency, forecast encompassing, is then tested in a multiple encompassing framework using the MS^* test statistic put forth by Harvey and Newbold (Equation 7). Finally, the sufficient condition for market

efficiency is also examined and compared to the results of both the traditional forecast evaluation and encompassing tests.

Empirical Results

Forecast Accuracy and Encompassing Tests

The traditional root mean squared error (RMSE) forecast accuracy measure is presented in Table 1, along with the mean absolute error (MAE) and Theil's U. At a one-quarter horizon (Panel A), the futures market records the smallest RMSE of 6.73%. The seasonal time series model (TS-1) has the largest RMSE of 17.96%. Similarly, both the MAE and Theil's U rank the futures market as the most accurate forecast at the one-quarter horizon. This result is consistent with previous findings of futures market pricing efficiency at short horizons (Irwin, Gerlow, and Liu). However, at two-quarters ahead (Table 1, Panel B), the USDA forecasts are the most accurate by all measures. The USDA's RMSE is 16.64% compared to 17.26% for the futures market, which is more accurate than either of the time series alternatives. Similarly, at a three-quarter horizon, the futures market does not have the smallest squared prediction error. The TS-1 and USDA have RMSE of 19.03% and 19.05%, respectively, while the futures market's RMSE is 19.63%. Therefore, based on the casual observation of these forecast accuracy measures at the two and three-quarter horizons, the milk futures market may meet the necessary condition to reject efficiency. However, it is important that the differences in prediction errors are statistically significant.

The MDM test in Equation 6 is used to test the statistical difference in MSE. The results for each forecast horizon are reported in Table 2. At the one-quarter horizon, the futures market clearly provides a superior forecast. The futures market has a statistically smaller mean squared

forecast error than each of the competitors (1% level). At this horizon, the USDA forecasts are statistically more accurate than either time series model (5% level), and there is no statistical difference between the time series forecasts. However, at two-quarters ahead (Panel B) and three-quarters ahead, the MDM test cannot distinguish between the accuracy of any of the forecasts. So, although the USDA and time series models produce smaller mean squared forecast errors than the futures based forecasts at these horizons (Table 1), they are not statistically smaller. This result would suggest that the futures market meets the traditional necessary condition for forecast efficiency because the other forecasts do not produce statistically smaller errors (Irwin, Gerlow, and Liu). However, as shown previously, this conclusion may be misleading if the futures forecast does not encompass all the information in the competing forecasts.

A more stringent test of pricing efficiency is forecast encompassing. That is, the futures market forecast should include all the information contained in alternative forecasts. The multiple forecast encompassing test shown in Equation (7) is conducted at each horizon using the futures market as the preferred forecast. The null hypothesis of market efficiency—that the futures market encompasses the competing forecasts—is tested with a F-test of MS^* in Equation (7). For the sake of completeness, the encompassing test is performed using each forecast as the preferred forecast (Harvey and Newbold).

The p-values from the encompassing tests are presented in Table 3. Using the futures market as the preferred forecast (first row of Table 3) represents the test for futures pricing efficiency. The null hypothesis that the futures price encompasses the information contained in the competing forecasts is rejected at the 1% level at the two two-quarter horizon. The null hypothesis is not rejected at conventional levels for the one- and three-quarter ahead forecasts.

The evidence suggests that at the two-quarter horizon, the futures market does not contain all the information in the competing forecasts. The encompassing necessary condition for market inefficiency is met. Importantly, this result is consistent with the USDA's smaller RMSE at this horizon. But, the encompassing test provides much stronger statistical evidence than the test for differences in MSE. That is, based on the traditional difference in MSE test (Table 2), the necessary condition to reject market efficiency is not met. Whereas, forecast encompassing is rejected (Table 3), which suggests that the necessary condition to reject efficiency is met. This result is consistent with Ashley, who reports that tests for differences in MSE may have low power against the null of equal forecast error variance. Therefore, the encompassing test is not only theoretically more appropriate, but it may also provide greater statistical power than tests for differences in MSE.

The remainder of the encompassing tests supports this finding. When the USDA and time series forecasts are designated as the preferred forecasts, in most instances they do not encompass the competing forecasts. The exception is at the two-quarter horizon, where forecast encompassing is not rejected for either the USDA or TS-1 forecasts, which confirms that they encompass the futures forecast at this horizon.

Collectively, the results suggest that the fluid milk futures market is efficient relative to forecasts produced by the USDA and univariate time series models at the one-quarter and three-quarter horizon. However, at the two-quarter horizon, the null hypothesis that the futures market contains all the information in the competing forecasts is rejected. This implies the potential for trading profits using the alternative forecasts at this horizon. Therefore, while the necessary condition for futures market inefficiency is met at this horizon, an examination of the sufficient

condition is needed to confirm the necessary condition and to make any definitive statements regarding market efficiency.

The Sufficient Condition for Market Inefficiency—Existence of Trading Profits

The sufficient condition for market inefficiency is the existence of net risk-adjusted trading profits. As pointed-out by Fama, this criterion is a joint test of excess returns and the risk pricing model. Before considering the appropriate risk model, it is necessary to test for statistically and economically significant profits. In other words, can using the alternative forecast procedures produce net positive trading returns?

To test for trading profits, a simple trading strategy is devised for the user of each alternative forecast: USDA, TS-1, and TS-2. If the alternative forecast is greater (less) than the futures forecast, then three futures contracts are bought (sold). The three contracts represent one for each delivery month in the quarterly price being forecast. The respective futures position is held until the next forecasting date, at which time the forecasts are re-evaluated versus the appropriate futures forecast. The position is either reversed or maintained, depending on the trade signal generated. In the final quarter (associated with the one-quarter horizon forecast), it is assumed that each contract is held to expiration. This allows the position to be effectively cash-settled to the quarterly average price. Transactions are assumed to occur at the daily settlement price on the day the *WASDE* reports are released.

As an example of this procedure, consider the forecast for the 2001.4 quarterly average price. On April 9, 2001 the October, November, and December futures averaged \$12.57 per hundredweight. The following morning, the *WASDE* report had a three-quarter ahead forecast of \$11.54.⁵ The signal is to sell one each of the October, November, and December futures. This

transaction is completed at the close of that day at an average price of \$12.63. Three months later, on July 10, the futures forecast for the fourth quarter is \$13.23, and the next day's USDA two-quarter ahead forecast is \$12.97. The trade signal is to maintain the short position. Finally, on October 10 the futures forecast is \$12.39, and on the morning of October 11 the USDA one-quarter ahead forecast is \$13.31. The trade signal in this case is to buy futures. That is, the short position is exited and a new long position is established at the closing price of \$12.45 on the close of the *WASDE* release day. The initial short position generated a gross gain of \$0.18 per hundredweight (\$12.63-\$12.45). The new long positions in the October, November, and December contracts are simply held until contract expiration when they are cash-settled. In this specific example, the cash settlement is at an average price of \$12.57, resulting in a gross gain of \$0.12 per hundredweight (12.57-12.45).

The trading revenue generated from this type of trading rule is reported in Table 4. Trading revenues are broken-out by forecast horizon. The temporal separation of the trading profits provides insight as to the nature of the revenue and its relationship to the forecast encompassing results. For instance, in the final three months (one-quarter horizon), trades based on the USDA forecasts generated a loss of \$0.264 per hundredweight or \$528 on a 200,000 pound contract. But, consistent with the forecast encompassing results, gross profits of \$0.272 per hundredweight are generated on trades held from six to three months prior to the quarter (two-quarter horizon). USDA trades held over the interval of nine to six months prior to expiration generated positive gross returns of \$0.070 per hundredweight. These results are very consistent with the forecast encompassing results (Table 3). That is, the futures forecast encompasses the competing forecasts at the one-quarter horizon; so, using the alternative forecasts to trade does not generate positive revenue. Conversely, because the futures market

does not incorporate all the information in the USDA forecasts at the two-quarter horizon, positive gross profits are available to traders using the USDA forecasts. The pattern of results across the horizons is similar for the other forecasting methods.

Certainly, the sufficient condition for market inefficiency is not met. Although the strategies produce positive mean revenue at the two-quarter horizon—as was suggested by the encompassing results—they are not statistically different from zero (t-test). Aggregated across horizons, revenue is close to zero. Moreover, the cost of trading these new, and often thin, markets in terms of both transaction costs and commissions would decrease the returns significantly. In this case, a further examination of the profits on a risk-adjusted basis is not warranted. Yet, this does not detract from the main point of conducting these tests. That is, the rejection of forecast encompassing for the futures market is associated with positive gross profits from using the competing forecasts. This adds some empirical support for evaluating pricing efficiency with a forecast encompassing approach.

Summary, Conclusions, and Discussion

This research shows that the traditional MSE necessary condition for pricing inefficiency in futures markets is not stringent enough. That is, a smaller MSE does not necessarily imply that a forecast is technically efficient relative to other forecasts. In light of this, the traditional MSE necessary condition may have low power against the null hypothesis of efficient markets. Here, we propose the multiple forecast encompassing test of Harvey and Newbold (2000) as a more appropriate test of pricing efficiency. That is, the futures market forecast should encompass all the information contained in competing forecasts.

The forecast encompassing test is applied to the CME's Class III milk futures market. Forecast horizons from one- to three-quarters ahead are examined. Simple time series models and USDA expert opinion generate the competing forecasts. Care is taken to map the USDA expert forecasts for "all milk" to the Class III price underlying the cash-settled futures; thereby, providing a direct test of forecast efficiency. The data utilized spans from July 1997 through June 2003. Since this futures contract cash settles to the USDA announced Class III price, any pricing uncertainty that may arise due to delivery options is avoided (Hranaiova and Tomek).

The fluid milk futures forecasts perform admirably at the one- and three-quarter horizons, producing the smallest MSE and encompassing the other forecasts. However, at the two-quarter horizon, the null hypothesis that the futures market contains all the information in the competing predictions is rejected. Therefore, the null hypothesis of pricing efficiency is rejected. A further examination of gross trading profits related to the competing forecasts at each horizon confirms the encompassing results. That is, positive gross trading revenues are generated by the competing forecasts at the two-quarter horizon. However, the returns are not statistically significant and of doubtful economic importance given trading costs.

In summary, this research presents a more stringent necessary condition for futures market efficiency than the traditional comparative analysis of mean squared forecast errors. Specifically, to reject the null of market efficiency, it is necessary that the futures market fails to encompass all competing forecasts. Further, the multiple forecast encompassing test of Harvey and Newbold (2000) provides a robust statistical test under a number of distributional assumptions. Given the market's tendency to quickly incorporate new information and eliminate inefficiencies (Timmerman and Granger), it seems unlikely that the fluid milk futures will continue to omit the information contained in alternative forecasts, like those produced by the

USDA. Furthermore, as the market develops, and trading costs decrease, the market will likely more fully incorporate additional information. Nonetheless, the presented methodology represents a step forward in refining and investigating the necessary conditions for futures market efficiency.

References

- Ashley, R. "Statistically Significant Forecasting Improvements: How Much Out-of-Sample Data is Likely Necessary?" *International Journal of Forecasting*. 19(2003): 229-239.
- Bessler, D.A. and J.A. Brandt. "An Analysis of Forecasts of Livestock Prices." *Journal of Economic Behavior and Organization*. 18(1992):249-263.
- Diebold, F.X. and R.S. Mariano. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*. 13(1995):253-263.
- Fama, E. "Efficient Capital Markets: II." *Journal of Finance*. 46(1991): 1575-1617.
- Garcia, P., R.M. Leuthold, T.R. Fortenbery, and G.F. Sarassoro. "Pricing Efficiency in the Live Cattle Futures Market: Further Interpretation and Measurement." *American Journal of Agricultural Economics*. 70(1988):162-169.
- Garcia, P. Hudson, M.A., and Waller, M.L. "The Pricing Efficiency of Agricultural Futures Markets: An Analysis of Previous Research Results." *Southern Journal of Agricultural Economics*. 20 (1988): 119-130.
- Gardner, B.L. "Futures Prices in Supply Analysis." *American Journal of Agricultural Economics*. 58(1976):81-85.
- Granger, C.W.J. "Can We Improve the Perceived Quality of Economic Forecasts?" *Journal of Applied Econometrics*. 11(1996):455-473.
- 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*. 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): 281-291.
- Harvey, D.I., S.J. Leybourne, and P. Newbold. "Tests for Forecast Encompassing." *Journal of Business and Economic Statistics*. 16(1998): 254-259.
- Harvey, D.I. and P. Newbold. "Tests for Multiple Forecast Encompassing." *Journal of Applied Econometrics*, 15(2000): 471-482.
- Hranaiova, J. and W.G. Tomek. "Role of Delivery Options in Basis Convergence." *Journal of Futures Markets*. 22(2002):783-809.
- Hurt, C. and P. Garcia. "The Impact of Price Risk on Sow Farrowings." *American Journal of*

- Agricultural Economics*. 64(1982):565-568.
- Irwin, S.H., M.E. Gerlow, and T. Liu. "The Forecasting Performance of Livestock Futures Prices: A Comparison to USDA Expert Predictions." *Journal of Futures Markets*. 14(1994):861-875.
- Just, R.E. and G.C. Rausser. "Commodity Price Forecasting with Large-Scale Econometric Models and the Futures Market." *American Journal of Agricultural Economics*. 63(1981):197-208.
- Kaminsky, G. and M.S. Kumar. "Efficiency in Commodity Futures Markets." *IMF Staff Papers*. 37(1990):670-699.
- Kenyon, D., E. Jones, and A. McGuirk. "Forecasting Performance of Corn and Soybean Harvest Futures Contracts." *American Journal of Agricultural Economics*. 75(1993):399-407.
- Leitch, G. and J.E. Tanner. "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures." *American Economic Review*. 81(1991): 580-590.
- Leuthold, R.M. "The Price Performance on the Futures Market of a Non-Storable Commodity—Live Beef Cattle." *American Journal of Agricultural Economics*. 56(1974):271-279.
- Leuthold, R.M., J.C. Junkus, and J.E. Cordier. *The Theory and Practice of Futures Markets*. Lexington, Massachusetts: Lexington Books, 1989.
- Leuthold, R.M., P. Garcia, B.D. Adam, and W.I. Park. "Examination of the Necessary and Sufficient Conditions for Market Efficiency: the Case of Hogs." *Applied Economics*. 21(1989):193-204.
- Leuthold, R.M. and P.A. Hartmann. "A Semi-Strong Form Evaluation of the Efficiency of the Hog Futures Market." *American Journal of Agricultural Economics*. 63(1979):482-489.
- Ma. C.W. "Forecasting Efficiency of Energy Futures Prices." *Journal of Futures Markets*. 9(1989):393-419.
- Martin, L. and P. Garcia. "The Price-Forecasting Performance of Futures Markets for Live Cattle and Hogs: A Disaggregated Analysis." *American Journal of Agricultural Economics*. 63(1981):209-215.
- Osborn, D.R., S. Heravi, and C.R. Birchenhall. "Seasonal Unit Roots and Forecasts of Two-Digit European Industrial Production." *International Journal of Forecasting*. 15(1999):27-47.
- Rausser, G.C. and C. Carter. "Futures Market Efficiency in the Soybean Complex." *Review of Economics and Statistics*. 65(1983):469-478.

Stein, J.L. "Speculative Price: Economic Welfare and the Idiot of Chance." *Review of Economics and Statistics*. 63(1981):223-232.

Timmerman, A. and C.W.J. Granger. "Efficient Market Hypothesis and Forecasting." *International Journal of Forecasting*. Forthcoming.

Tomek, W.G. "Commodity Futures Prices as Forecasts." *Review of Agricultural Economics*. 19(1997):23-44.

Table 1. Accuracy Measures

Panel A: One-Quarter Ahead				
	Futures	USDA	TS-1	TS-2
RMSE	0.0673	0.1113	0.1796	0.1578
MAE	0.0591	0.0946	0.1515	0.1318
Theil's U	0.2722	0.4502	0.7262	0.6380
Panel B: Two-Quarter Ahead				
	Futures	USDA	TS-1	TS-2
RMSE	0.1726	0.1664	0.2018	0.2016
MAE	0.1465	0.1322	0.1746	0.1636
Theil's U	0.6956	0.6707	0.8133	0.8123
Panel C: Three-Quarter Ahead				
	Futures	USDA	TS-1	TS-2
RMSE	0.1963	0.1905	0.1903	0.2144
MAE	0.1696	0.1464	0.1573	0.1679
Theil's U	0.7575	0.7353	0.7345	0.8277

Table 2. MDM Test for Difference in Mean Squared Prediction Errors

Panel A: One-Quarter Ahead			
	USDA	TS-1	TS-2
Futures	0.0048 ^a	0.0009	0.0027
USDA		0.0081	0.0174
TS-1			0.2999
Panel B: Two-Quarters Ahead			
	USDA	TS-1	TS-2
Futures	0.8108	0.3768	0.4285
USDA		0.3407	0.3167
TS-1			0.9947
Panel C: Three-Quarters Ahead			
	USDA	TS-1	TS-2
Futures	0.8825	0.8202	0.6432
USDA		0.9962	0.6942
TS-1			0.5757

^aP-value from the MDM test for difference in mean squared errors.

Table 3. Test for Multiple Forecast Encompassing

Preferred Forecast	Forecast Horizon		
	One-Ahead	Two-Ahead	Three-ahead
Futures	0.1712 ^a	0.0040	0.1500
USDA	0.0242	0.9973	0.0252
TS-1	0.0055	0.9999	0.0024
TS-2	0.0042	0.0002	0.2540

^aP-value for the null that the preferred forecast encompasses all the competing forecasts.

Table 4. Trading Revenue in First Quarter of Horizon, \$/cwt.

Forecasts	Forecast Horizon			
	One-Ahead	Two-Ahead	Three-ahead	Total
USDA	-0.264 (0.159) ^a	0.272 (0.309)	0.070 (0.153)	0.057 (0.441)
TS-1	-0.287 (0.157)	0.317 (0.307)	-0.016 (0.154)	-0.025 (0.304)
TS-2	-0.412 (0.144)	0.456 (0.299)	0.171 (0.154)	0.0400 (0.376)

^aStandard error in parenthesis.

Endnotes

¹ The sufficient condition for market inefficiency is the ability to produce risk-adjusted trading profits.

² It is important to distinguish that prior studies, such as Leuthold, Garcia, Adam, and Park, simulate the past with out-of-sample forecasts; but, they are *ex post* in the sense that they are not made in real-time. In contrast, extension or export forecasts—such as those examined by Irwin, Gerlow, and Liu—are made in real-time. Thus, a test using a history of these forecasts is truly *ex ante*.

³ The terms “preferred” forecast and “competing” forecast are commonly used in the forecast encompassing literature, and do not relate to any *a priori* judgment of the performance of the forecasts.

⁴ Class III and BFP milk prices are available starting in May of 1995. Prior to that, the “all milk” price is used to calculate a return series. The two return series are combined to make a continuous series of log-relative price changes.

⁵ The \$11.54 is derived using the USDA “all milk” forecast of \$12.40 incorporating the estimated mapping relationship discussed above.