Selected paper for presentation at AAEA Annual Meeting, Chicago, August 2001

New MSE Tests for Evaluating Forecasting Performance: Empirics & Bootstrap

By

Carlos W. Robledo

Email: wrobledo@agro.uncor.edu

Hector O. Zapata

Email: hzapata@agctr.lsu.edu

and

Michael McCracken*

Email:mmccrac@unix1.sncc.lsu.edu

Abstract: Two asymptotically valid out-of-sample MSE tests have been developed by Diebold-Mariano (1995) and Stock-Watson (1999). The empirical usefulness of the tests is illustrated through a U.S. wheat model estimated with fixed, recursive and rolling forecasting schemes. Bootstrap methods are adopted to reflect small sample size effect on tests..

---

* Robledo is a Ph.D. candidate and Zapata a professor, Department of Agricultural Economics and Agribusiness, Louisiana State University Agricultural Center; Michael McCracken is an assistant professor, Department of Economics, Louisiana State University, Baton Rouge, Louisiana.

Copyright 2001 by Sung C. No and Hector O. Zapata. 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.

New MSE Tests for Evaluating Forecasting Performance: Empirics & Bootstrap

Carlos W. Robledo, Hector O. Zapata, and Michael W. McCracken

Introduction

Recent changes in U.S. agricultural policy have encouraged the development of a more market-oriented agricultural sector with fewer barriers to trade. Examples of such changes in policy are the Federal Agriculture Improvement and Reform Act of 1996 (FAIR Act) and the signing or ratification of international agreements, such as the North American Free Trade Agreement (NAFTA) and the World Trade Organization (WTO). A unique feature of the Fair Act is that it decoupled production decisions from price subsidies; as a result, agricultural commodity markets are expected to be more sensitive to changes in market fundamentals. Under these new conditions, profitability and survival in the industry depend on obtaining market information that accurately reflects market trends. Farmers and agribusinesses that trade in grain markets, for instance, use *outlook information* for deciding when and how to buy or sell. Policy makers usually need *projections* of production and prices to design better policies that affect the financial health of the industry. This challenge will continue through, and may be exacerbated by the expiration of the FAIR Act in the year 2002. Therefore, the evaluation of the forecasting ability of the models used to construct useful economic forecasts becomes an issue that calls for a renewed attention.

In this study we examine the forecasting ability of two competing models for the U.S. wheat market. We chose to work with a U.S. wheat market model because its structure is well documented in the literature (Chambers and Just) and is simple in nature when compared to those of other agricultural commodities, such as corn and soybeans.

Chambers and Just's model specification is simple and approximates the wheat market through linkages between the U.S. macro economy, export market through exchange rates, and U.S. commodity markets. The model is used to forecast U.S. wheat domestic consumption, inventories, exports, prices, and production. A variation of this model that uses the Kansas City wheat prices, instead of those in Chicago, will be considered the alternative and competing model. This respecification is considered relevant because Kansas is a large wheat producing state, with a large share of the U.S. market. Much of the wheat produced in the U.S. is also exported. Thus prices in Kansas reflect transactions for a large share of the domestic and international markets.

The evaluation of forecasting accuracy via measures of point estimates is a well-established practice in the forecasting literature. The mean error (ME), the error variance (EV), the mean square error (MSE), and the mean absolute error (MAE) are often used for evaluating forecasting performance. The usual practice for choosing among alternative forecasting models has been to select a model that shows a lower accuracy measure, but with no attempt in general to asses its sampling uncertainty. In this sense, Parks (1990), Kastens (1996), and Dorfman (1998) are good exceptions. The sampling uncertainty of point estimates of forecast accuracy has recently received considerable attention in the econometric and forecasting literature. In Section 3 we describe a test the null hypothesis of no difference in the accuracy of two competing forecasts for the U.S. wheat market using the Diebold and Mariano (1995) test and the test suggested by Stock and Watson (1999). Details on the form of these tests and their statistical properties are also provided, following McCracken and West (1999). Section 4 a bootstrap technique for postsample model validation (Ashley) is used to quantify the uncertainty in median inference due to small sample size. We generate forecasts using three different schemes: fixed, recursive, and rolling schemes (McCracken and West, 1999). Section 5 summarizes the main results, and a brief summary concludes the paper.

The illustration used in the forecasting experiment

The reference-forecasting model

The reference-forecasting model adopted in this paper is a VAR system derived from a dynamic econometric model having four equations (production, disappearance, inventories and exports) and an identity that clears the markets for a given commodity (Chambers and Just, 1981). This specification incorporates predetermined endogenous variables (lagged variables) to model the dynamics of the system and links the U.S. macro economy with the U.S. commodity market by means of the exchange rate, which is an exogenous variable that explains U.S. exports.

The basic structural model adopted is as follows:

(1) Disappearance: $PWD_t = \beta_{1,0}+\beta_{1,1}PWD_{t-1}+\beta_{1,2}RPDI_t+\beta_{1,3}FALL_t+\beta_{1,4}WINT_t + \beta_{1,5}SPRI_t+e_{1,t}$

(2) Inventory: $PWI_t = \beta_{2,0}+\beta_{2,1}PWI_{t-1}+\beta_{2,2}RPW_t+\beta_{2,3}FALL_t+\beta_{2,4}WINT_t+\beta_{2,5}SPRI_t+e_{2,t}$

(3) Exports: $PWX_t = \beta_{3,0}+\beta_{3,1}PWX_{t-1}+\beta_{3,2}RPW_t+\beta_{3,3}SDR_t+\beta_{3,4}THPW_t+\beta_{3,5}FALL_t+\beta_{3,6}WINT_t+\beta_{3,7}SPRI_t+e_{3,t}$

(4) Production: $PWPR_t = \beta_{4,0}+\beta_{4,1}RWAP_{t1}+\beta_{4,2}RWSP_t+e_{4,t}$

(5) Identity: $PWPR_t + PWI_{t-1} = PWD_t + PWI_t + PWX_t$,

where PWD = per capita US wheat disappearance (bushels per person),

RPDI = real per capita disposable income (chained by Wholesale Price Index- Base 1992),

FALL, WINT, SPRI = dummy variables for the respective seasons of the year,

PWI = per capita US wheat inventories (bushels per person),

RPW = real wheat price (No. 2 SRW at Chicago),

PWX = per capita US wheat exports (bushels per person),

SDR = exchange rate (SDR per dollar),

THPW = European Union threshold price of wheat (units of account per metric ton),

PWPR = per capita US wheat production (bushels per person),

RWAP = real average price of U.S. wheat received by farmers,

RWSP = real support price of U.S. wheat,

$e_{1,t}, \ldots, e_{4,t}$ = random error terms independent and identically distributed.

To complete the information needed to estimate the model, the series PWD, PWI, PWX, RPW, and PWPR are assumed to be endogenous, while every other variable is assumed to be exogenous, as suggested by economic theory.

The above structure is used as the base information set from which the following a vector autoregressive (VAR) specification, the forecasting model adopted in the forecasting evalution, is obtained:

(6) $$\mathbf{y}_{1,t+1} = \boldsymbol{\mu}_1 + A_{1,1}\mathbf{y}_{1,t} + \ldots + A_{1,k}\mathbf{y}_{1,t-k} + \boldsymbol{\Psi}_1\mathbf{D}_{t+1} + \mathbf{u}_{1,t+1},$$

where $\mathbf{y}_{1,t+1}$ represents a (5×1) vector consisting of the endogenous variables PWD_{t+1}, PWI_{t+1}, PWX_{t+1}, RPW_{t+1}, and $PWPR_{t+1}$, and the sub-index 1 identifies the reference model; $\boldsymbol{\mu}_1$ is a vector of constants; $\mathbf{A}_{1,1}$ through $\mathbf{A}_{1,k}$ are (5×5) parameter matrices associated with the first k lags of the endogenous vector $\mathbf{y}_{1,\ t+1}$; $\mathbf{\Psi}_1$ is a (5×8) parameter matrix associated with the vector $\mathbf{D}_{t+1}$ that contains the exogenous variables[1] $\Delta RPDI_t$, ΔSDR_t, $RWAP_t$, $\Delta THPW_{t+1}$, $\Delta WSTOSKCW_t$, $\Delta RWSP_{t+1}$, $FALL_{t+1}$, $WINT_{t+1}$, and $SPRI_{t+1}$; $\mathbf{u}_{1,\ t+1}$ is an error vector (5×1) at time t+1.

The Data

The U.S. wheat data used are from the USDA-ERS Situation and Outlook Yearbook 1999 report and the USDA-ERS World Agricultural Demand and Supply Estimates (WASDE-BB), while the U.S. macroeconomic data are obtained from CITIBASE. The frequency of the data available is quarterly, for the period 1981:3-1999:4. Table 1 provides summary statistics for the endogenous variables. For brevity, the results on unit root tests based on the augmented Dickey-Fuller and Phillips-Perron tests are not presented in the paper. Since most of the variables significantly show to have a unit root and that cointegration is present (Engle and Granger, 1987), the following error correction model is adopted,

$$(7)\qquad \Delta\mathbf{y}_{1,\ t+1} = \boldsymbol{\mu}_1 + \boldsymbol{\Gamma}_{1,\ 1}\ \Delta\mathbf{y}_t + \ldots + \boldsymbol{\Gamma}_{1,\ k-1}\ \Delta\mathbf{y}_{(t-k+1)+1} + \boldsymbol{\pi}_1\ \mathbf{y}_{1,t} + \boldsymbol{\Psi}_1\mathbf{D}_{t+1} + \mathbf{u}_{1,\ t+1}$$

where Δ represents the difference operator, $\boldsymbol{\pi}_1 = \alpha_1\phi_1'$, α_1 is a loading matrix, ϕ_1 contains the cointegrating vectors, and all the other terms in (7) are as in (6).

Table 1

Summary Statistics of Endogenous Variables (1975:3-1998:4)

NAME	N	MEAN	ST. DEV	VARIANCE	MINIMUM	MAXIMUM
A. Endogenous variables						
PWD	94	10.809	4.9185	24.191	5.3606	24.207
PWI	94	71.434	31.817	1012.3	14.209	139.20
PWX	94	13.042	3.1554	9.9568	6.7255	21.084
RPW	94	3.5075	0.97047	0.94182	2.0995	6.4318
PWPR	94	24.231	42.046	1767.9	0.00000	121.23
B. Exogenous variables						
RPDI	94	16697.	1792.1	0.32116E+07	13447.	19829.
SDR	94	122.46	17.475	305.37	96.260	174.62
THPW	94	227.58	112.11	12568.	0.00000	352.99
RWAP	94	0.86739	1.5605	2.4351	0.00000	6.1592
RWSP	94	0.86947	1.6202	2.6252	0.00000	4.3800

[1] Differences are used (Δ operator) for those variables that are nonstationary and assumed exogenous. Note also that $\Delta RPDI_t$, ΔSDR_t, $\Delta WSTOCKW_t$, $\Delta RWAP_t$ in $\mathbf{D}_{t+1}$ represent expectations for these changes in variables at period t+1

The alternative-forecasting model

The alternative-forecasting model emerges as a result of a challenging question posed on the economic relevance of using Chicago prices (No. 2 soft red winter wheat) to represent the endogenous variable U.S. wheat price in the forecasting equations given in (6) and (7). For example, Chambers and Just (1981) used this data without providing further justifications, although in the USDA-ERS Situation and Outlook Yearbook 1999 Report wheat prices for other U.S. markets such as Kansas City, Toledo, Saint Louis, Portland, Minneapolis, and the Gulf Ports in Texas are available. Recent studies used Gulf Port prices and Great Lake Port prices of wheat (see Mohanty et al., 1996a), for providing empirical evidence of relationships between U.S. and Canadian wheat prices, based on a cointegration analysis, but again, without providing any economic justification about the selection of those time-series. Mohanty et al. (1996b) provide an extensive discussion on the different studies related with the World wheat market, for which prices for the U.S. wheat were included, studies that again did not explain why the time-series data used to represent the U.S. wheat prices were selected. In synthesis, as the menu of time-series already available to model the U.S. wheat prices offers different alternatives, and following Meese and Rogoff (1983, 1988), one should select that time-series price that has a higher predictive power for the other variables in the model.

Which price series should be used in this VAR specification? Equation (5) models production as a function of farm price. As production is an endogenous variable, and the system in (6) is cointegrated (see Table 2), it is expected that farm prices better contribute to forecasting Kansas City price than do Chicago prices. If this is the case, we expect inventories will be better forecasted using Kansas City than Chicago prices, because inventories are a function of the U.S. wheat prices. On the contrary, as long as Chicago is a market more immediately affected by international conditions, then Chicago prices should outperform Kansas City prices in forecasting Exports. Related with U.S. wheat disappearance, although it is not a function of U.S. wheat prices, it is expected that if Kansas City prices for wheat are better forecasted, then disappearance should be better forecasted by the model using Kansas City prices. Therefore, by selecting Kansas City prices (No. 1, soft read winter) as an alternative time series for wheat prices, we are able to specify an alternative forecasting model to equation (6).

---

based on the information available at period t.

Table 2
Lag-Length Selection and Number of Cointegrating vectors (r)

	MODEL 1			MODEL 2		
LAG	SC	HQ	r	SC	HQ	r
1	3.0422	1.5344	3	3.1493	1.6395	3
2	3.5472	1.5317	2	3.7058	1.6903	1
3	3.4367	1.8748	1	4.5864	2.0245	0
4	5.2275	2.0587	3	5.2599	2.0912	2

Given these considerations, the alternative-forecasting model adopted is,

(8) $$\mathbf{y}_{2,\,t+1} = \boldsymbol{\mu}_2 + \mathbf{A}_{2,\,1}\mathbf{y}_{1,t} + \ldots + \mathbf{A}_{2,\,k}\mathbf{y}_{2,t-k} + \boldsymbol{\Psi}_1\mathbf{D}_{t+1} + \mathbf{u}_{2,\,t+1}\ ,$$

where $\mathbf{y}_{2,\,t+1}$ represents a (5×1) vector having the endogenous variables PWD, PWI, PWX, RPWK (real wheat price, No. 1 HRW at Kansas City), and PWPR at time t+1. The sub-index 2 identifies the alternative model, $\boldsymbol{\mu}_2$ is a (5×1) vector of constants for the alternative model, $\mathbf{A}_{2,\,1}$ through $\mathbf{A}_{2,\,k}$ are (5×5) parameter matrices for the first k lags of the endogenous vector $y_{2,\,t+1}$, $\boldsymbol{\Psi}_1$ is a (5×8) parameter matrix associated with the vector D_{t+1} which contains the exogenous variables ΔRPDI, ΔSDR, ΔTHPW, RWAP, ΔWSTOCKW, ΔRWSP, FALL, WINT, and SPRI as in the case of the reference model, and $\mathbf{u}_{2,\,t+1}$ is a (5×1) error vector at time t+1.

The associated error correction model to that in (8) is,

(9) $$\Delta\mathbf{y}_{2,\,t+1} = \boldsymbol{\mu}_2 + \boldsymbol{\Gamma}_{2,\,1}\,\Delta\mathbf{y}_{2,\,t} + \ldots + \boldsymbol{\Gamma}_{2,\,k-1}\,\Delta\mathbf{y}_{2,(\,t-k+1)+1} + \boldsymbol{\pi}_2\, y_{2,\,t} + \boldsymbol{\Psi}_2\mathbf{D}_{t+1} + \mathbf{u}_{2,\,t+1}$$

where Δ represents the difference operator, $\boldsymbol{\pi}_2 = \alpha_2\phi_2'$, α_2 is a loading matrix, ϕ_2 contains the cointegrating vectors, and all the other terms in (9) are as in (8).

Testing equality of forecast ability

As mentioned earlier, we are interested in forecasting U.S. wheat disappearance, inventories, exports, prices, and production. We have already proposed two models that can be used to predict these variables (i.e. the non-stationary models in equation (6) and (8), or their stationary counterparts in equation (7) and (9)). The question we are interested in is "which model is better and how to make that decision?"

Using the selected models, and splitting the sample into in-sample (t=1,…,R) and out-of-sample (t=R+1,…,T) portions, we can construct one-step ahead forecasts of the form $\hat{y}_{1,t+1}$ and

$\hat{y}_{2,t+1}$, from time t=R,…,T−1. Hence we observe the P=T−R+1 forecasts errors $\{\hat{u}_{1,t+1}, \hat{u}_{2,t+1}\}_{t=R}^{T-1}$, where

$$\text{(10)} \qquad \hat{u}_{1,t+1} = y_{1,\,t+1} - \hat{y}_{1,t+1} \,,$$

and

$$\text{(11)} \qquad \hat{u}_{2,t+1} = y_{2,\,t+1} - \hat{y}_{2,t+1} \,.$$

To determine whether one forecasting model is more accurate than the other, and letting $u^2_{1,t+1}$ and $u^2_{2,t+1}$ be the square of $u_{1,t+1}$ and $u_{2,t+1}$ defined in (6) and (8), respectively[2], we test the following hypothesis,

$$\text{(12a)} \qquad H_{0,DM}: Eu^2_{1,t+1} - Eu^2_{2,t+1} = 0 \quad \text{vs.} \quad H_{1,DM}: Eu^2_{1,t+1} - Eu^2_{2,t+1} \neq 0\,,$$

$$\text{(12b)} \qquad H_{0,SWa}: \frac{Eu^2_{1,t+1}}{Eu^2_{2,t+1}} - 1 = 0 \quad \text{vs.} \quad H_{1,SWa}: \frac{Eu^2_{1,t+1}}{Eu^2_{2,t+1}} - 1 \neq 0, \text{ and}$$

$$\text{(12c)} \qquad H_{0,SWb}: \frac{Eu^2_{2,t+1}}{Eu^2_{1,t+1}} - 1 = 0 \quad \text{vs.} \quad H_{1,SWb}: \frac{Eu^2_{2,t+1}}{Eu^2_{1,t+1}} - 1 \neq 0.$$

McCracken and West (1999) provide an extensive review of papers that provide asymptotic and finite-sample evidence of tests for the null in (12). Of special interest is that of

---

[2] Note that $\hat{u}_{1,t+1}$ and $\hat{u}_{2,t+1}$ are (5×1) vectors, with scalar entries $\{\hat{u}_{1,i,t+1}\}$ and $\{\hat{u}_{2,i,t+1}\}$, i=1, …, 5. In what follows ,and for the remainder of the paper, the sub-index i will be dropped in the notation of the scalars $\hat{u}_{1,i,t+1}$, and $\hat{u}_{2,i,t+1}$. Thus $\hat{u}_{1,t+1}$ and $\hat{u}_{2,t+1}$ are scalars now and thereafter.

Corradi, Swanson, and Olivetti (1999), where they outlined conditions under which the both tests test for predictive ability can be extended to the case of two forecasting models, each of which may include cointegrating relations, when allowing for parameter estimation error.

Forecasting schemes

Typically, the out-of-sample comparison is made in two stages. First, the forecasts of interest are constructed using the reference model, and then a second time using the alternative model. Second, with the two sequences of forecast errors at hand, tests of equal forecast accuracy are conducted. Before presenting the two tests considered in this paper we first describe the three forecasting schemes adopted, following West and McCracken (1998). These are the *fixed*, *recursive*, and *rolling* schemes.

Under the *fixed* scheme, each forecast is generated using parameter estimates $\hat{\beta}_{i,R}$, i=1,2, obtained only once, using the in-sample data (t=1,…,R). Hence each forecast $\hat{y}_{i,t+1}(\hat{\beta}_{i,R}) \equiv \hat{y}_{i,t+1}$, t=R,…,T−1, depends upon the same parameter estimate.

Under the *recursive* scheme, the parameter estimates $\hat{\beta}_{i,t}$ are obtained with added data as forecasting moves forward through time. The first prediction, $\hat{y}_{i,R+1}(\hat{\beta}_{i,R}) \equiv \hat{y}_{i,R+1}$, is created using the parameter estimates based on data from 1 to R. The second prediction, $\hat{y}_{i,R+2}(\hat{\beta}_{i,R+1}) \equiv \hat{y}_{i,R+2}$ is created using the parameter estimates obtained on data from 1 to R+1, and so on. In general, for t=R,…,T, the prediction $\hat{y}_{i,t+1}(\hat{\beta}_{i,t+1}) \equiv \hat{y}_{i,t+1}$ is created using the parameter estimates based on data from 1 to t.

Under the rolling scheme, forecasts are constructed for t=R,…,T using the most recent R observations[3]. The first prediction is created using the parameter estimates based on data from 1 to R. The second prediction is created using the parameter estimates obtained on data from 2 to R+1, and so on. In general, for t=R,…,T, the prediction yt+1 is created using the parameter estimates based on data from t–R+1 to t.

---

[3] Note that for the fixed and rolling schemes the parameter estimates also depend on the choice of R. Rather than introduce another subscript we leave that notation implicit.

Out-of-sample tests of forecast accuracy

West (1996) proved that in the environment provided by (10), (11), and (12),

$$(13)\qquad P^{1/2}\left[P^{-1}\sum_{t=R}^{T-1}\left(\begin{pmatrix}\hat{u}^2_{1,\ t+1}\\ \hat{u}^2_{2,\ t+1}\end{pmatrix}-\begin{pmatrix}\sigma^2_1\\ \sigma^2_2\end{pmatrix}\right)\right]\xrightarrow{d} N(0,V)$$

where the long-run covariance matrix V is positive definite and depends upon whether or not the squared forecast errors are serially correlated (i.e. ARCH-type behavior[4]).

To test the null in (12a) Diebold and Mariano (1995) use the statistic,

$$(14)\qquad DM=\frac{P^{-1/2}\sum_{t=R}^{T-1}\hat{d}_{t+1}}{[\hat{V}_1]^{1/2}}\xrightarrow{d} N(0,1),$$

where

$$(15)\qquad \hat{d}_{t+1}=\hat{u}^2_{1,t+1}-\hat{u}^2_{2,t+1}=\gamma'_1\begin{pmatrix}\hat{u}^2_{1,t+1}\\ \hat{u}^2_{2,t+1}\end{pmatrix},\ \gamma_1=(1,-1)',\ \text{and}$$

$$(16)\qquad \hat{V}_1 \text{ is a consistent estimator of } V_1=\gamma'_1 V\gamma_1.$$

For the case of no ARCH behavior we chose the following consistent estimator of V, following the results exposed in West and McCracken (1998),

$$(16a)\qquad \hat{V}=P^{-1}(X'X)^{-1}X'\mathbf{11}'X(X'X)^{-1},$$

where $X=(X_1,X_2)$ is a $(P\times2)$ matrix, $X_i=\hat{u}^2_i-MSE_i$ $(i=1,2)$, MSE_i is the associated mean squared error defined elsewhere, and $\mathbf{1}$ is a $(P\times1)$ vector of ones.

For the case of ARCH-type behavior we choose the Newey-West (1987) serial correlation consistent estimator version of (16a).

To test the null hypothesis in (12b) Stock and Watson (1999) use the statistic,

$$(17) \qquad SWa = \frac{P^{1/2}}{\hat{V}_{2a}^{1/2}}\left[\frac{P^{-1}\sum_{t=R}^{T-1}\hat{u}_{1,t+1}^{2}}{P^{-1}\sum_{t=R}^{T-1}\hat{u}_{2,t+1}^{2}} - 1\right] \xrightarrow{d} N(0,1),$$

where $\hat{V}_{2a}$ is a consistent estimator of $V_{2a} = \gamma_2' V \gamma_2$, $\gamma_2 = (MSE_2^{-1}, -MSE_1/MSE_2^2)'$, and V is estimated accordingly as in (16a) or its serial correlation consistent version, for the no ARCH or ARCH case, respectively.

Alternatively, to test the null hypothesis in (12c), the reverse of the statistic in (17) is used, i.e.,

$$(18) \qquad SWb = \frac{P^{1/2}}{\hat{V}_{2b}^{1/2}}\left[\frac{P^{-1}\sum_{t=R}^{T-1}\hat{u}_{2,t+1}^{2}}{P^{-1}\sum_{t=R}^{T-1}\hat{u}_{1,t+1}^{2}} - 1\right] \xrightarrow{d} N(0,1),$$

where $\hat{V}_{2b}$ is a consistent estimator of $V_{2b} = \gamma_3' V \gamma_3$, $\gamma_3 = (MSE_1^{-1}, -MSE_2/MSE_1^2)'$, and V as before.

An empirical evaluation using these asymptotic results was reported in Robledo et al (2000).

Double Bootstrap Postsample Evaluation

Ashley (2000) introduces a new technique for assessing whether one sequence of postsample forecasting errors is smaller than another. The appeal of this techniques lies in the fact that postsample forecasting performance is usually constrained by small sample size, and this condition renders existing testing procedures (e.g., Diebold and Mariano) valid only in large samples. The superiority of results using this technique is documented in their simulation and empirical results.

---

[4] See Harvey, Leybourne, and Newbold (1997) for a discussion of tests of forecast accuracy in the presence of ARCH.

A typical question in forecasting experiments is whether the forecast generated from one model are superior (in a MSE sense) to those of a contender. Ashley develops a bootstrap-based inference approach by testing the hypothesis of whether the expected size of the postsample forecast errors from one model significantly exceeds that of another, based on an observed sequence of N postsample forecasting errors from each model. The approach is summarized as follows:

1. Take the observed sequences of postsample forecast errors from two alternative models and specify a bivariate VAR structure that adequately represents the two sequences. This model is a descriptive parameterization of the serial correlation structure of the forecast errors from the two alternative models. No information about the models that generated these forecast errors is needed.
2. Choose a relative forecast performance evaluation criteria, and use this criteria, say r, to calculate the probability that r is less than or equal to some given value, c (usually c=1). The population value of r could be given by MSE_2/MSE_1, where 1 and 2 denote the two competing models (other choices are given in Ashley). This version of the ratio r is consistent with the Stock-Watson criterion.
3. Test the null hypothesis Ho: r <= c against the alternative hypothesis Ha: r > c based on the observed forecast error sequences from models 1 and 2 (alternatively Ho: r >= c vs. Ha: r <c).
4. Use the bootstrap procedures to calculate the sampling error factor distributions that quantify the inferential uncertainty caused by sampling errors in the estimation of the VAR model for the error sequences and by the bootstrap approximation.

Main Results

The forecast errors for each of the VAR series in the two models (using Kansas City prices vs. Chicago prices) are used for illustration purposes. In order to provide the forecasts and to perform the tests of predictive ability, the in-sample size R was chosen to be 1981:3-1995:4. As previously mentioned, the FAIR Act was passed in 1996 and hence it is of interest to study the forecast ability of the competing models after this more market oriented policy was enacted. The optimal lag-length selected is k=2 for the reference ECM model given in equation (7), and k=2 for the alternative ECM model of equation (9). OLS is used to estimate the parameters needed to construct the forecasts under the different schemes, as long as consistent estimates are only needed. Empirical results for equations (12b)-(12.c) are reported in Robledo et at (2000).

The MSE from both models, asymptotic and bootstrap p-values are summarized in Table 3. The sections are divided into the three forecasting schemes: fixed, recursive and rolling. Error sequences for each of the variables in the VAR were generated and the mean-squared errors calculated. The main findings can be summarized as follows.

- The asymptotic p-values for a MSE test that one model forecasts an ith variable in a VAR better than an alternative model generates, in general, mixed results by forecasting scheme: fixed, recursive, and rolling. One exception is in forecasting disappearance: using the Kansas City wheat price is superior to that of using Chicago's price and all three forecasting schemes suggest it.
- Fixed Forecasting Scheme: there is considerable discrepancy between the asymptotic p-value versus the bootstrap one. For U.S. wheat disappearance and, wheat inventories and production, the conclusion from the asymptotic p-value favors the forecasts generated from the model that uses Kansas City prices. The bootstrap p-values favors this model for two of these variables: disappearance and production.
- Recursive Forecasting Scheme: When forecasting recursively, the asymptotic and bootstrap p-values lead to the same conclusion at the .05 level, the Kansas City price model works better in forecasting disappearance only. Note however, that there is considerable discrepancy in the magnitude of the asymptotic p-value and that of the bootstrap.
- Rolling Forecasting Scheme: In this case, the bootstrap approach would suggest that the Kansas City price model would forecast better for disappearance, inventories, and production; the asymptotic p-value approach would reveal the same but for the first two variables only.

VI. Summary

This paper applies a recently developed bootstrap technique to assess the forecasting performance of alternative forecasting models. The illustration used focuses on the role of the prices at spatially separated markets for U.S. wheat, namely Kansas City and Chicago, in the estimation of a multiple time series (VAR) model. This preliminary evaluation of postsample forecasting performance points to considerable discrepancies in the magnitude of estimated asymptotic p-values when compared to p-values generated from the bootstrap technique. These discrepancies carry over various forecasting schemes, namely fixed, recursive and rolling forecasting. Although much more work is needed in assessing the usefulness of this approach in evaluating forecasting performance, it is clear that the use of asymptotic p-values generated from recently developed postsample techniques should be carefully reexamined. Perhaps one area of research where the validity of this bootstrap approach may be of relevance is in work dealing with causal relationships and other dynamic responses.

Table 3. Stock-Watson Test of Forecasting Performance, Asymptotic and Bootstrap Significance, U. S. Wheat Model, Selected Forecasting Schemes, 16 observations.

Variable	MSE1	MSE2	Asymptotic p-value	Bootstrap p-value	
	FIXED SCHEME				
PWD	26.34	19.25	<0.001	0.0450	
PWI	355.72	325.41	0.0478	0.1595	*
PWX	13.54	15.39	0.2069	0.9490	
RPW	0.19	0.17	0.8674	0.4300	
PWPR	133.64	120.99	0.0086	0.0255	
	RECURSIVE SCHEME				
PWD	18.66	15.25	<0.0001	0.0500	
PWI	276.50	264.17	0.1395	0.2160	
PWX	7.5435	8.8288	0.09500	0.9620	
RPW	0.1350	0.1098	0.74819	0.3915	
PWPR	118.70	112.78	0.26586	0.1055	
	ROLLING SCHEME				
PWD	18.45	14.85	<0.0001	0.0185	
PWI	269.18	251.78	<0.0001	0.0895	*
PWX	7.73	8.90	0.1593	0.8810	
RPW	0.15	0.11	0.6875	0.3520	
PWPR	116.10	109.02	0.3294	0.0300	*

Hypothesis tested: Ho: MSE2/MSE1>=1 vs.H1: MSE2/MSE1< 1.

Stock-Watson ratio=MSE2/MSE1.

References

Chambers, R.G. and R.E. Just. (1981): "Effects of Exchange Rate Changes on U.S. Agriculture: A Dynamic Analysis," *American Journal of Agricultural Economics*, 1, 32-46.

Clark, T. E. , and M. W. McCracken (1999): "Test of Equal Forecast Accuracy and Encompassing for Nested Models", Working Paper, Federal Reserve Bank of Kansas City and Louisiana State University.

Corradi, V., N. R. Swanson, and C. Olivetti (1999): "Predictive Ability With Cointegrated Variables," Working Paper, Department of Economics, Texas A&M University.

Diebold, F. X. (1988): *Elements of Forecasting*. South-Western College Publishing.

Diebold, F. X. and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.

Dorfman, J.H. (1998): "Bayesian Composite Qualitative Forecasting: Hog Prices again," *American Journal of Agricultural Economics*: 80(3):543-552.

Engle, R. F. and C. W. J. Granger (1987): "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251-276.

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

Kastens, A.L. (1997) " Forecasting retail food prices under current conditions," *American Journal of Agricultural Economics*: 78(2):301-313.

Kenyon, D; J. Eluned, and A. McGuirk (1993) " Forecasting performance of corn and soybean harvest futures contracts," *American Journal of Agricultural Economics*: 75(2):399.

Knutson, R.D., J.B. Penn, and B.L. Flinchbaugh (1998): *Agricultural and Food Policy*. Prentice Hall Ed, New Jersey, U.S.

Lütkephol, H. (1993): *Introduction to Multiple Time Series Analysis,* Second Edition. Springer-Verlag Ed., Germany.

McCracken, M. W. (1999): "Asymptotics for Out of Sample Tests of Causality," Working Paper, Department of Economics, Louisiana State University.

McCracken, M. W., and K. D. West (1999): "Inference About Predictive Ability," Working Paper, University of Wisconsin and Louisiana State University.

Meese, R. A., and K. Rogoff (1983): "Empirical Exchange Rate Models of the Seventies: Do they Fit Out of Sample?," *Journal of International Economics*, 14, 3-24.

Meese, R. A., and K. Rogoff (1988): "Was it Real? The Exchange Rate –Interest Rate Differential Over the Modern Flotaing Rate Period," *Journal of Finance*, 43, 933-948.

Mohanty, S., D. B. Smith, and W. F. Peterson (1996a): "Time Series Evidence of Relationships Between U.S. and Canadian Wheat Prices," Working Paper 96-WP154, Center for Agricultural and Rural Development, Iowa State University.

Mohanty, S., W. H. Myers, and D. B. Smith (1996b): "A Reexamination of Price Dynamics in the International Wheat Market," Working Paper 96-WP71, Center for Agricultural and Rural Development, Iowa State University.

Park, T. (1990) " Forecast Evaluation for Multivariate Time-Seires Models: The U.S. Cattle Market," *Western Journal of Agricultual Economics*, 15(1):133-143.

Robledo, C.W., H. O. Zapata, and M. Mccracken. "The Predicted Ability of Two U.S. Wheat Market Models." Selected Paper, AAEA Annual Meeting 2000, Tampa, Florida, August 2000.

Spriggs, J. M. K., and D. Bessler (1982): "The lead-Lag Relationship Between Canadian and U.S. Wheat Prices," *American Journal of Agricultural Economics*, 64, 569-572.

Stock, J.H. and M.W. Watson (1999): "Forecasting Inflation," Working Paper 7023, National Bureau of Economic Research, http://www.nber.org/papers/w7203.

West, K. D. (1996): "Asymptotic Inference About Predictive Ability," *Econometrica*, 64, 1067-1084.

West, K.D., and M. W. McCracken (1998): "Regression-Based Tests of Predictive Ability," *International Economic Review*, 39(4): 817-840.