Testing Unbiasedness in Futures Markets: A Clarification

Scott E. Hein
Christopher K. Ma
S. Scott MacDonald

I. INTRODUCTION

Recently, discussions have been raised with respect to the validity of the unbiasedness tests in futures markets. For empirical testing, researchers usually estimate some variant of the following equation to test whether futures prices are unbiased predictors of future spot prices:

$$S_{i,m} = a + bF_{i,m-n} + v_{i,m} \tag{1}$$

where $S_{i,m}$ is the spot price at maturity time, m, for the i-th futures contract, and $F_{i,m-n}$ is the futures price of the i-th contract n days prior to maturity time, m. The empirical test of the unbiasedness involves the use of an OLS estimate of the eq. (1) to perform an F test on the joint null hypothesis: $a = 0$ and $b = 1$. The collective evidence from previous studies suggest that $a > 0$ and $b < 1$; i.e., the unbiasedness hypothesis is rejected in the futures markets.

In a recent article of this Journal, Maberly (1985) questions the validity of an OLS procedure estimating eq. (1). Maberly argues that the OLS estimates are biased as a result of an *ex-post* censored data problem related to the forecast error. *Ex-post* the forecast error and the forecast are negatively correlated. More recently, Elam and Dixon (1988) disagree with this characterization of the source of the problem. They argue that if the spot/futures price changes follow a random walk, eq. (1) turns in a regression model with a lagged dependent variable and it is then this lagged dependent variable which has caused an excessively higher number of rejections of the unbiasedness hypothesis in previous studies.

Based on Elam and Dixon (1988), if the spot price is described by the following nonstationary process,

$$S_{i,t} = S_{i,t-1} + -u_t \tag{2}$$

Scott E. Hein is the First National Bank at Lubbock Distinguished Scholar and Area Coordinator at Texas Tech University.

Christopher K. Ma is an Associate Professor at Texas Tech University.

S. Scott MacDonald is an Associate Professor at Texas Tech University.

The Journal of Futures Markets, Vol. 10, No. 5, 555–562 (1990)

CCC 0270-7314/90/050555-08$04.00

where u_t is an i.i.d. random error, the spot price at time m can be shown as:

$$S_{i,m} = S_{i,m-n} + u_{m-n} + u_{m-n+1} + \ldots + u_m$$

Under the assumption that both current spot prices and futures prices are unbiased predictors of future spot prices; i.e., $S_{i,m-n} = E(S_{i,m}) = F_{i,m-n}$, the testing of eq. (1) actually turns into estimating the following lagged dependent variable regression:

$$S_{i,m} = a + bS_{i,m-n} + v_{i,m} \tag{3}$$

where

$$v_{i,m} = u_{m-n} + u_{m-n+1} + \ldots + u_m.$$

Elam and Dixon argue (page 368):

"... the OLS estimates for a and b in eq. (3) are biased because the explanatory variable is a lagged value of the dependent variable. The bias is caused because there is correlation between the independent variable and the previous values for the error term (Johnston (1972, pp. 304–06) and Judge (1985, pp. 166–167))...."

Elam and Dixon provide simulations of futures and spot prices as prescribed by the above unbiased assumption and under these simulations, they estimate eq. (1). The estimation results offer evidence that there is a higher than usual number of F tests rejecting the unbiasedness hypothesis when, in fact, the unbiasedness hypothesis is true.

While Elam and Dixon have brought insight into the unbiasedness testing methodology, their arguments leave two key questions unanswered. First, even in the presence of a lagged dependent variable as an explanatory variable, Johnston (1972) shows that the OLS estimates of parameters are biased only when *both* the lagged dependent variable is included as an explanatory variable *and* the random error v_t is serial correlated[1]. Note that in eq. (3), since u_t is i.i.d., the sum of u_t, v_t is i.i.d., there is not *a priori* reason to suspect that the v_t are serial correlated. For this point, Elam and Dixon provide no explanation as to why the errors, $v_{i,m-n}$, in eq. (3) and in their simulations might be serial correlated and thus pose an estimation problem. This leaves the puzzling question why their simulations still locate the rejection bias. Second, in the estimation results they provide evidence of vastly different rejection rates for forecast horizons of 13 weeks or longer versus shorter forecast

[1]Actually, a more fundamental problem, as mentioned by a number of previous studies, is that the validity of F tests is seriously complicated by the nonstationarity of the time series. Dickey and Fuller (1979) show that there is a bias on the estimates of the time series with unit root such as the spot price specified in eq. (2). Engel and Granger (1987) further argue that since nonstationary variables have infinite variances which make the t tests or F tests invalid, the standard hypothesis testing does not apply to economic series with unit roots. This posits a serious problem especially for the unbiasedness testing in eq. (3) or eq. (4) since spot and futures prices are either assumed to have unit roots (Maberly (1985) Elam and Dixon (1988)), or found to be nonstationary (Ma and Hein (1990)). It may be presumptuous to attribute the excessive rejections of the unbiasedness hypothesis to nonstationarity without further evidence, it does suggest that the previous findings on unbiasedness are under serious suspicion if the issue of (non)stationarity of the variables involved in the testing is not first resolved.

horizons. Only in the longer forecast horizon setting are there the sharp over-rejection rate they argue exists. But, why does the length of the forecast horizon matter?

THE ESTIMATION PROBLEM

The purpose of this article is to answer both of these questions. The identity of one source of serial correlations is detailed in Elam and Dixon's own setting, and, perhaps, in previous studies. The problem is due to a specific type of overlapping data. As argued above, the overlapping data problem cannot be explicitly observed from eq. (3), since the specification clearly requires that the parameters are estimated over a cross-sectional sample (versus continuous time series) of different contracts ($i = 1, k$). Most researchers, including Elam and Dixon, recognize the bias on parameters estimates from using overlapping time series, so they use only one observation with the identical time to maturity per contract to test equation (3). However, it is important to note that when the futures prices are observed from a distant time to maturity, such that there is a forecasting horizon equal to or longer than the delivery cycle of the contract, the testing of eq. (3) will unavoidably suffer from autocorrelated residuals.

The nature of the overlapping data problem in the Elam-Dixon setting can be demonstrated by explicitly writing out the "i-th" observation (contract) in estimat-

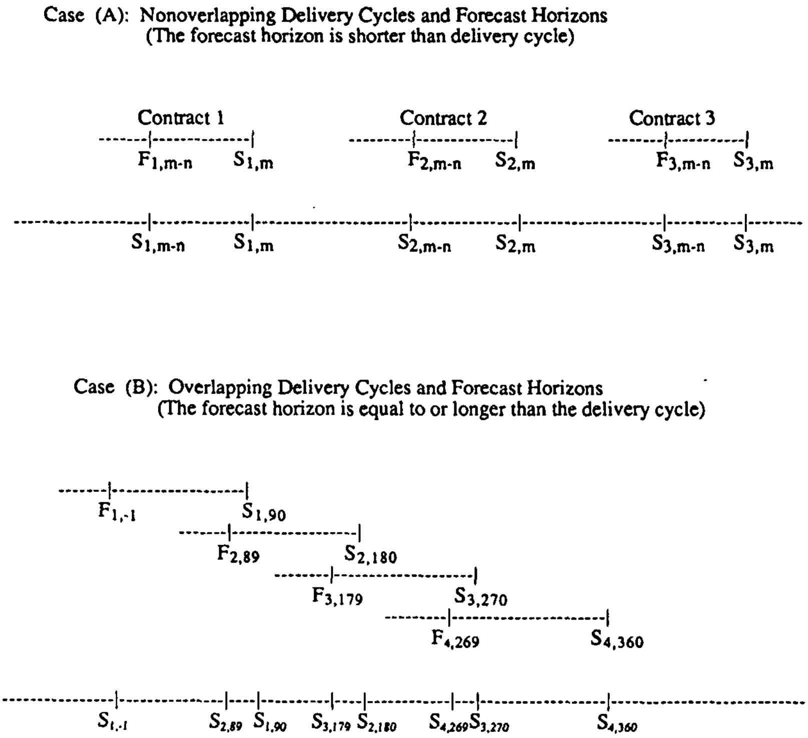

Figure 1
Delivery cycles of different futures contracts in the sample.

ing eq. (3):

$$S_{i,m} = a + bS_{i,m-n} + u_{m-n} + u_{m-n+1} + \ldots + u_m \qquad \text{for } i = 1,\ldots k.$$

where k is the number of different contracts in the sample, m is the length of the delivery cycle, and n is the length of the forecasting horizon (observation period). To show the impact of the relationship between m and n, two contrasting examples are illustrated in Figure 1. In Case A when the forecasting horizon is shorter than the delivery cycle of the futures contract, each pair of the three observations, $(S_{1,m-n}, S_{1,m})$, $(S_{2,m-n}, S_{2,m})$, $(S_{3,m-n}, S_{3,m})$ on the time series are clearly independent because they do not share any common terms in residuals.

On the other hand, in Case (B) where the forecasting horizon is assumed to be, say 91 days, which exceeds the assumed delivery cycle of the previous futures contract of 90 days, there are always two common days which are shared by two neighboring contracts. From eq. (3), each observation can be written then as:

$$S_{2,180} = a + bS_{2,89} + v_{2,180}$$

$$S_{3,270} = a + bS_{3,179} + v_{3,270}$$

$$S_{4,360} = a + bS_{4,269} + v_{4,360}$$

Consider the first two residuals, $v_{2,180}$, $v_{3,270}$, from eq. (3),

$$v_{2,180} = u_{180} + u_{179} + \ldots\ldots\ldots + u_{89}$$

$$v_{3,270} = u_{270} + u_{269} + \ldots\ldots\ldots + u_{179}$$

The covariance of these two residuals can be represented as:

$$\begin{aligned}
&\mathrm{Cov}(v_{2,180}, v_{3,270})\\
&\quad = \mathrm{Cov}(u_{180} + u_{179} + \ldots + u_{89}, u_{270} + u_{269} + \ldots + u_{179})\\
&\quad = E[(u_{180} + u_{179} + \ldots + u_{89})(u_{270} + u_{269} + \ldots + u_{179})]\\
&\qquad - E(u_{180} + u_{179} + \ldots + u_{89})E(u_{270} + u_{269} + \ldots + u_{179}).
\end{aligned}$$

Since u_t is mean zero and i.i.d.,

$$\begin{aligned}
&= E[(u_{180} + u_{179})(u_{180} + u_{179})]\\
&= E(u_{180}^2 + u_{179}^2) > 0 \qquad (4)
\end{aligned}$$

Expression (4) clearly shows that the error terms are not independent, but suffers positive first-order autocorrelations. That is, whenever the forecasting horizon is longer than the delivery cycle by one day, there is positive serial correlation between residuals of eq. (3). Also, the magnitude of the serial correlation is greater if the length of the overlapping period is larger.[2] Equation (4) explains why there is serial correlation in the residuals of eq. (1) as estimated by Elam and Dixon, (at least when the forecast horizon is long). It also explains why Elam and Dixon find the "false" rejection rates to be higher for samples with longer time from maturity.

A careful examination of Elam and Dixon's simulation results which are summarized in Table I provides evidence consistent with the concern about overlapping

[2]For example, for the case that the forecast horizon overlaps with the delivery cycle by two days, the covariance term, $\mathrm{Cov}(v_{2,180}, v_{3,270})$, between residuals becomes $E(u_{180}^2 + u_{179}^2 + u_{178}^2)$.

Table I
SUMMARY OF ELAM AND DIXON'S SIMULATIONS
(TAKEN FROM TABLE II ON PAGE 370)

Time to Maturity	No. of Significant F Test* for $a = 0$ and $b = 1$	No. of Significant F Tests (25 Samples)***
1 week	0**	10
3 weeks	0	7
6 weeks	1	20
13 weeks	6	47
20 weeks	6	84
24 weeks	6	97
% Rejection Rate		
1–6 weeks	5.55%	8.22%
13–24 weeks	100%	50.67%

*The F test is conducted over the all (25) subsamples replicated with 6 different sample sizes (26, 50, 100, 200, 500, 1000 observations in each subsample).

**0 suggests that 0 out of 6 samples has significant F value.

***The F test is conducted on each of the 25 subsamples replicated with 6 different sample sizes. Total number of F tests is 150 for each time to maturity.

data. Note that when using the significant F tests on the entire 25 samples for every specific time to maturity, the rejection rate for samples with time to maturity less than 6 weeks is 5.55 percent (one significance out of eighteen F tests, a reasonable and acceptable level for Type I error[3]). However, when the time to maturity of the samples go beyond 12 weeks which is the length of the delivery cycle used in the simulation, the rejection rate jumps to 100 percent (all eighteen F tests are rejected). Since Elam and Dixon, following Maberly, assume that the delivery cycle is 12 weeks, most of what they find as "false" rejections concentrate on the samples which have longer forecasting horizon (13–24 weeks).

AN EMPIRICAL EXPOSITION USING REAL WORLD SAMPLES

To further show that the validity of these arguments go beyond a mere simulation artifact, some empirical evidence is obtained from Ma and Hein (1990) who test the unbiasedness hypothesis over more that 30 different types of futures contracts. In their study, eq. (3) is estimated using actual futures prices and spot prices. The same estimation is replicated over samples with different forecast horizons. The relevant information of one example, the T-Bill futures contracts with the delivery cycles of 90 days, is summarized in Table II. The results are striking! Most of the rejections of the unbiasedness hypothesis concentrate on samples with time to maturity longer than 90 days. More importantly for the very long horizon, along with

[3]Having considered the impact of the overlapping data problem on the F tests, the higher percentages of rejection rates when Elam and Dixon conduct the F tests on each of the 25 subsamples replicated over 6 different sample sizes remain bothersome. The higher-than-usual numbers of significant F values are demonstrated in the second column of Table I. Note that the rejection rates for the samples with time to maturity less than 12 weeks are still considered too high (8.22%, or 37 out of 450 cases). This implies that the overlapping data problem may not be the only issue, and the nonstationarity of the time series may have complicated the validity of the test statistics. To correct for this problem, see Ma and Hein (1990).

Table II
UNBIASEDNESS TESTS IN T-BILL FUTURES CONTRACTS
(TAKEN FROM MA AND HEIN (1990))

No. of Days Forecasts	No. of Obs.	*a* estimate (Std. error)	*b* estimate (Std. Error)	F tests $a = 0, b = 1$	Durbin Watson
20	48	0.1328 (0.128)	0.9695 (0.028)	0.9240	1.803
30	49	−0.0541 (0.208)	1.0121 (0.046)	0.1837	2.370
40	52	0.0253 (0.225)	0.9942 (0.049)	0.0240	2.275
50	57	0.0401 (0.269)	0.9912 (0.059)	0.0916	2.209
70	55	0.4347 (0.412)	0.9034 (0.091)	0.7431	2.049
90	51	1.0378 (0.475)	0.7703 (0.105)	2.3992	1.956
100	50	1.0630 (0.521)	0.7650 (0.115)	2.3002	1.817
110	50	0.7659 (0.481)	0.8305 (0.106)	1.2844	1.942
120	55	0.6489 (0.479)	0.8566 (0.106)	1.0937	1.738
130	52	−5.5635 (1.324)	2.2266 (0.292)	9.3328	1.146
140	55	0.7433 (0.479)	0.8359 (0.106)	1.6335	1.388
160	48	0.8629 (0.561)	0.8089 (0.124)	1.1912	1.248
170	45	1.1317 (0.593)	0.7495 (0.1314)	1.8956	1.649
180	46	1.6809 (0.645)	0.6281 (0.142)	3.5342	1.206
190	53	1.5370 (0.590)	0.6601 (0.130)	3.6916	1.201
210	55	1.011 (0.624)	0.7567 (0.138)	1.8695	1.239
230	51	1.2979 (0.649)	0.7129 (0.143)	2.1547	1.079
240	50	1.6549 (0.641)	0.6338 (0.142)	3.4449	1.030
250	47	1.4618 (0.688)	0.6764 (0.152)	2.2836	0.849
260	49	1.4502 (0.619)	0.6792 (0.137)	3.0192	0.877
280	52	1.3126 (0.621)	0.7100 (0.137)	2.8454	0.656
300	42	1.1636 (0.689)	0.7430 (0.152)	2.0433	0.872

Table II
UNBIASEDNESS TESTS IN T-BILL FUTURES CONTRACTS (TAKEN FROM MA AND HEIN (1990)) continued

310	44	1.2646	0.7201	1.5590	0.733
		(0.754)	(0.167)		
320	49	1.4376	0.6821	2.3194	0.607
		(0.721)	(0.159)		
330	51	1.8569	0.5892	3.8695	0.693
		(0.709)	(0.157)		
350	49	2.3109	0.4884	5.4627	0.614
		(0.719)	(0.159)		
370	47	2.3814	0.4728	5.2686	0.760
		(0.766)	(0.169)		
380	39	1.9535	0.5675	2.7157	1.097
		(0.903)	(0.200)		
390	44	203350	0.4831	4.4179	0.816
		(0.814)	(0.180)		
400	47	2.3568	0.4783	4.5412	0.703
		(0.839)	(0.186)		
420	48	2.5418	0.4374	5.8081	0.622
		(0.802)	(0.177)		

the most significant F values, the Durbin-Watson statistics are significant and suggest that the residuals are positively serial correlated (See Fig. 2). This is completely consistent with predictions.

In summary, this note provides a possible reason for the large numbers of F-tests rejecting the unbiasedness hypothesis in previous studies. It appears that the failure

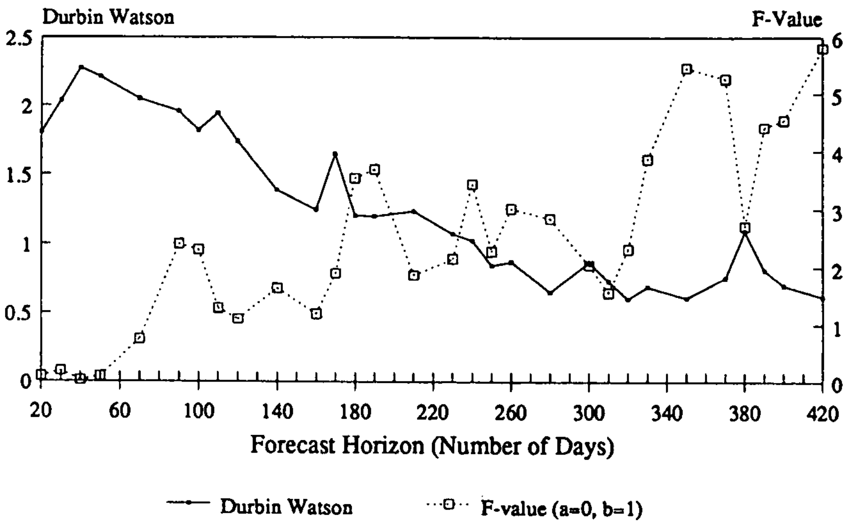

Figure 2
Durbin Watson and F-values by forecast horizon.

to control the serial correlation, which resulted from the overlap between the delivery cycle and the forecast horizon, may generate biased estimates as well as misleading conclusions.

Bibliography

Dickey, D. A., and Fuller, W. A. (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association,* 74:427–431.

Elam, E., and Dixon, B. L. (1988): "Examining the Validity of A Test of Futures Market Efficiency," *Journal of Futures Markets,* 8:365–372.

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

Ma, C., and Hein, S. (1990): "Testing Unbiasedness in Futures Markets: Further Evidence," Working Paper, Texas Tech University.

Maberly, E. D. (1985): "Testing Futures Market Efficiency—A Restatement," *Journal of Futures Markets,* 5:425–432.