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

Prequential Approach to Turning Point Prediction

by

Heon Kook and David A. Besler

Suggested citation format:

Kook, H., and D. A. Besler. 1992. "Prequential Approach to Turning Point Prediction." Proceedings of the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. Chicago, IL. [http://www.farmdoc.uiuc.edu/nccc134].

Prequential Approach to Turning Point Prediction

Heon Kook and David A. Bessler[1]

Dawid (1984) considers forecasts to be probability distributions over future events. He proposes the "prequential" approach to data analysis, which derives from the terms "prediction" in a "sequential" manner. The prequential approach is founded on the premise that a prime purpose of statistical analysis is to make sequential probability predictions on future observations. He argues that it is important to focus on models which do a good job in forecasting observables (the next data point) rather than focus on unobservables (parameters). Prequential forecasting involves making a forecast, observing the outcome of an event, and formulating forecasts of subsequent events, drawing on accumulating information.

Two techniques are commonly used for the prediction of turning points: one is signal detection from leading indicators and the other is stochastic simulation of future events. The former includes the works of Zarnowitz and Moore (1982), Neftci (1982), and Diebold and Rudebusch (1987). The latter includes the works of Wecker (1979), Kling (1987), and Zellner et al. (1991). As Kling pointed out, there are two problems with the applications of the leading-indicator approach. First, it generally does not predict the date on which the turning point is to occur but simply implies that a turning point is imminent. Second, it does not provide an explicit probability statement expressing the forecaster's (model's) degree of certainty regarding an upcoming turning point.

The purpose of this study is to apply prequential analysis to turning point prediction on a set of daily agricultural prices. The paper is presented in the following sections. First, we define our model of turning point prediction. The model is essentially that defined earlier by Kling (1987). Our second section considers metrics on probability forecast evaluation. Here we argue that previously applied metrics (the chi-squared goodness of fit statistic and the probability score) offer limited "ranking power" on assessments of ordered events (which our turning point forecasts are an example). We suggest the ranked probability score (Epstein, 1969; Murphy, 1970) be used for these evaluations. Our third section is an application of prequential analysis to turning point forecasts of three cotton price series. Data sources and the statistical model used to generate the probability forecasts are described. Evaluation of these forecasts is also presented here. The final section of the paper offers conclusions and suggestions for further research.

Turning Point Prediction

A turning point is labeled as either a peak or a trough. A peak is defined as the point which is greater than or equal to τ preceding values and τ subsequent values and a

---

[1] Heon Kook is an economist with the Korea Food Research Institute. David A. Bessler is a professor of Agricultural Economics at Texas A&M University. This research was funded in part by Texas Higher Education Commission ARP project 117520. Texas Agricultural Experiment Station Paper No. 30587.

trough is defined as the point which is less than or equal to τ preceding values and τ subsequent values. We are interested in the date at which a turning point will occur. As this date is uncertain, we express our beliefs (model's beliefs) in terms of probabilities. In other words, we are interested in assigning a probability to successive dates that each will be the next peak or trough. If we have predictive distributions up to k steps ahead at time t, we can predict turning points which occur from $t+1-\tau$ to $t+k-\tau$.

Define two indicator variables z_{it}^{P} and z_{it}^{T} such that

$$(1) \qquad z_{it}^{P} = \begin{cases} 1 & \text{if a turning point peak occured at time t for series i;} \\ 0 & \text{otherwise,} \end{cases}$$

and

$$(2) \qquad z_{it}^{T} = \begin{cases} 1 & \text{if a turning point trough occured at time t for series i;} \\ 0 & \text{otherwise.} \end{cases}$$

Stated alternatively:

$$(3a) \qquad z_{it}^{P} = \begin{cases} 1 & \text{if } x_{it} \geq x_{i,t+k},\ k = -\tau,-\tau+1,\ldots,-1,1,\ldots,\tau-1,\tau; \\ 0 & \text{otherwise,} \end{cases}$$

and

$$(3b) \qquad z_{it}^{T} = \begin{cases} 1 & \text{if } x_{it} \leq x_{i,t+k},\ k = -\tau,-\tau+1,\ldots,-1,1,\ldots,\tau-1,\tau; \\ 0 & \text{otherwise.} \end{cases}$$

The future values of the random time series X_t and the two indicator variables z_{it}^{P} and z_{it}^{T} can be characterized in terms of predictive distribution functions; that is,

$$(4) \qquad g_{t+1,\ldots,t+k}(x_{t+1},\ldots,x_{t+k} \mid x_t,x_{t-1},\ldots,x_1),$$

$$(5a) \qquad f^{P}_{t+1-\tau,\ldots,t+k-\tau}(z^{P}_{t+1-\tau},\ldots,z^{P}_{t+k-\tau} \mid x_t,x_{t-1},\ldots,x_1),$$

$$(5b) \qquad f^{T}_{t+1-\tau,\ldots,t+k-\tau}(z^{T}_{t+1-\tau},\ldots,z^{T}_{t+k-\tau} \mid x_t,x_{t-1},\ldots,x_1).$$

The predictions are probability distributions over the events $W_t^{P} = (W_{1t}^{P},\ldots,W_{mt}^{P})'$ ("time until the next turning-point peak") and $W_t^{T} = (W_{1t}^{T},\ldots,W_{mt}^{T})'$ ("time until the next turning-point trough"). These vectors are random and are defined by

$$(6a) \qquad w_{it}^{P} = w(z^{P}_{i,t+1-\tau},\ldots,z^{P}_{i,t+k-\tau}) = k,\ i = 1,\ldots,m,$$

where k is such that $z_{i,t+k}^{P} = 1$ and $z_{i,t+j}^{P} = 0$, for $j < k$; and

(6b) $w_{it}^{T} = w(z_{i,t+1-\tau}^{T},\ldots,z_{i,t+k-\tau}^{T}) = k,\ i = 1,\ldots,m,$

where k is such that $z_{i,t+k}^{T} = 1$ and $z_{i,t+j}^{T} = 0$, for $j < k$. Notice that W_{it}^{P} and W_{it}^{T}, $i = 1,\ldots,m$, have discrete probability distributions over the set of integers $\{-\tau+1,-\tau+2,\ldots,-1,0,1,2,\ldots,-\tau+k\}$. Thus, at time n, the forecaster with known values x_t, $t = 1,\ldots,n$, issues probability distributions P_{in}^{P} and P_{in}^{T} for the random events W_{in}^{P} and W_{in}^{T}, respectively.

Probability Evaluation Methods

The prequential principle, introduced by Dawid (1984), is that a model should be rejected only on the basis of the forecasts that it actually makes. A model should not be rejected *a priori* because it fails some preconceived criterion of goodness. In particular, one should focus on out-of-sample forecasts and ultimate realizations rather than on tests of model fit or *a priori* notions of model adequacy. Below we consider several metrics which have been used to evaluate probability forecasts. The first of these, sharpness (or standard deviation of the issued distribution), does not meet the prequential principle and thus is not recommended for use (solus) in probability forecast evaluation. The other measures considered are consistent with the prequential principle.

Sharpness (Murphy, 1973) measures the extent to which the probability forecast approaches the categorical forecast of one state among N states. Sharpness of probability forecasts measures the closeness to possible realization value, 0 or 1. Sharpness is defined as:

(7) $$SHP(p) = (1/T)\sum_{k=1}^{K} T_k p^k (u - p^k)',$$

where u is a unity vector ($u=[1,\ldots,1]$), T is the number of probability forecasts, T_k is the number of probability forecasts of distinct type k, where K is the number of distinct probability forecasts p^k (note, $\sum_{k=1}^{K} T_k = T$). The measure is defined on the interval $[0,(N-1)/N]$ where N = number of states with SHP = 0 corresponding to perfection.

Sharpness is closely related to the standard deviation of the distribution. The relationship between the standard deviation and sharpness can be illustrated by an example. Suppose two normal distributions $N_1(0,\sigma^2)$, and $N_2(0,4\sigma^2)$. The standard deviation of N_2 is two times greater than that of N_1. That is, N_1 is a tighter distribution than N_2. Divide the X space (outcome space) into six intervals such as $[-\infty,-2\sigma]$, $[-2\sigma,-\sigma]$, $[-\sigma,0]$, $[0,+\sigma]$, $[+\sigma,+2\sigma]$, and $[+2\sigma,+\infty]$. From the normal table, we can calculate the probabilities assigned to each interval:

p_1 = [0.0228 0.1359 0.3413 0.3413 0.1359 0.0228] and
p_2 = [0.1587 0.1498 0.1915 0.1915 0.1498 0.1587].

The sharpness measure for N_1 is 0.7286 while the sharpness of N_2 is 0.8314. We can see that N_1 is sharper than N_2. Bessler and Kling (1990) use the standard deviation to choose among two distributions (a preference was held for the distribution having the lower standard deviation), given that both are well-calibrated (defined below). As

sharpness and standard deviation do not involve the actual realized event, they do not meet the prequential principle.

Probability calibration requires issued probabilities to agree with their expost relative frequency. So that if a probability forecast of .25 is issued 100 times, calibraiton requires that, after the fact, we should observe 25 of these events to actually occur. Forecasts can be judged with respect to calibration by applying the probability integral transform to the observed fractiles from the cumulative distributions on each variable at each forecast point. Observed fractiles should be uniformly distributed for a set of well-calibrated distributions. A chi-squared goodness-of-fit statistic may be used to formally test the appropriateness of a particular set of observed outcomes (Dawid, 1984).

Kling pointed out that the empirical assessment of the probability distributions P_n^P and P_n^T is not reliable if the number of occurrences is small. For this reason he suggested a discrete event approximation of the probability integral transform. The cumulative distribution function F_{it}^P is calculated as:

$$(8) \qquad F_{it}^P(q) = \sum_{j=-\tau+1}^{q} P_{it}^P(j),$$

where $P_{it}^P(j)$ is the probability of a peak in month j, and q is the last month by which they are forecasted. Let $U_{it}^P = F_{it}^P(W_{it}^P)$, then U_{it}^P is a random variable taking values on the interval [0,1]. An estimated cumulative distribution function $F_U(U_{it}^P)$ for U_{it}^P is obtained by taking the observed sequence $u_{it}^P = F_{it}^P(w_{it}^P)$, t = 1,...,n; arranging the sequence in the order $u_i^P(1),...,u_i^P(n)$ of increasing value; and calculating

$$(9) \qquad F_U(u_i^P(j)) = \frac{j}{n}, \; j = 1,2,...,n.$$

This is called a calibration function. For a well-calibrated prequential forecasting system (PFS) the graph of the calibration function $F_U(u_i^P(j))$ should not be significantly different from a forty-five degree line. For a well-calibrated PFS, $u_i^P(1),...,u_i^P(n)$ should be uniformly distributed and chi-squared test can be applied to test that the PFS is well-calibrated.

Several scoring rules have been introduced and, as they each have a calibration component, can be used as evaluation metrics that are consistent with the prequential principle. The Brier score, or probability score, is perhaps the most commonly applied rule (Murphy 1972a, 1972b, 1973; DeGroot and Fienberg 1982, 1986; Winkler 1986). Consider a variable whose range has been divided into N mutually exclusive and collectively exhaustive states $\{s_1, ..., s_N\}$. Let the row vector $p_t = (p_{1t}, ..., p_{Nt})$ $(p_{nt} \geq 0, \Sigma_n p_{nt} = 1; n = 1,....,N; t = 1,...,T)$ denote the t forecast in a collection of T forecasts, where p_{nt} is the forecast probability of state s_n on the t^{th} forecast, and let the row vector $d_t = (d_{1t},...,d_{Nt})$ denote the t^{th} observation in the collection of T relevant observations, where d_{nt} equals one if state s_n obtains on the t^{th} observation and zero otherwise. Then PS for the collection of T forecasts $p_t(t=1,...,T)$ is PS(p,d), where

$$(10) \qquad PS(p,d) = (1/T) \sum_{t=11}^{T} (p_t - d_t)(p_t - d_t)'$$

in which a prime denotes a column vector. The PS is a probability forecast analog of mean-squared error (MSE) of point forecast. The range of the probability score is [0,2] with PS = 0 corresponding to perfect accuracy.

A problem which has been identified with the Brier score is that it doesn't recognize (or reward) closeness of probabilities to the state which actually obtains in multi-state cases. Epstein and Murphy proposed the Ranked Probability Scoring rule (RPS) which has more "ordering power" among multi-state probability forecasts than does the Brier score (Epstein 1969; Murphy 1970, 1972c). The RPS can rank probability forecasts which have the same Brier score. As Murphy argued, the PS is a particularly appropriate measure for the evaluation of forecasts of unordered variables, while the ranked probability score (RPS) is a particularly appropriate measure for the evaluation of forecasts of ordered variables (Murphy 1970). Unordered variables are variables in which the distance between values, or states is not meaningful, while ordered variables are variables in which the distance is meaningful. Questions such as which is taller the Eifel Tower or Big Ben or which is longer the Suez canal or Panama canal are unordered variables; whereas, probabilities on various price intervals are ordered variables. For continuous variables like prices, the distance between values is meaningful. Let the row vector $p_t = (p_{1t},..., p_{Nt})$ $(p_{nt} \geq 0, \Sigma_n p_{nt} = 1;$ $n = 1,....,N;$ $t = 1,...,T)$ denote the t forecast in a collection of T forecasts, where p_{nt} is the forecast probability of state s_n on the t^{th} forecast, and let the row vector $d_t = (d_{1t},...,d_{Nt})$ denote the t^{th} observation in the collection of T relevant observations, where d_{nt} equals one if state s_n obtains on the t^{th} observation and zero otherwise. To calculate the RPS, we need to define the cumulative row vector forecast $P_t = (P_{1t},...,P_{Nt})$ where

$$(11) \qquad P_{nt} = \sum_{m=1}^{n} p_{mt}$$

$(n = 1,...,N)$ and the cumulative observation row vector $D_t = (D_{1t},...,D_{Nt})$ where

$$(12) \qquad D_{nt} = \sum_{m=1}^{n} d_{mt}$$

$(n = 1,...,N)$. Then RPS for the collection of T forecasts p_t $(t=1,...,T)$ is RPS(p,d), where

$$(13) \qquad RPS(p,d) = \frac{1}{T} \sum_{t=1}^{T} (P_t - D_t)(P_t - D_t)'$$

in which a prime denotes a column vector. Note that the range of RPS(r,d) in equation (13) is the closed interval [0, N-1], where N is the number of states, 0 is the perfect score and N-1 is the worst score.

Suppose that we have two vector forecasts such as p^1 = (0.0, 0.2, 0.3, 0.4, 0.1, 0.0) and p^2 = (0.2, 0.3, 0.0, 0.4, 0.0, 0.1) and an observation vector d = (0, 0, 0, 1, 0, 0). Then the cumulative forecasts and observation are P^1 = (0.0, 0.2, 0.5, 0.9, 1.0, 1,0), P^2 = (0.2, 0.5, 0.5, 0.9, 0.9, 1.0), and D = (0, 0, 0, 1, 1, 1), respectively. The PSs of p^1 and p^2 are both equal to 0.5, even though p^1 has probabilities "closer" to the

state actually occurred. The RPS can rank (choose) between the two vectors. The RPS of p^1 is 0.3 while the RPS of p^2 is 0.56. So p^1 is the preferred forecast.

One can construct examples in which the probability score (PS) and the ranked probability score (RPS) give conflicting results. Suppose that we have two vector forecasts such as p^1 = (0.0, 0.1, 0.4, 0.4, 0.1, 0.0) and p^2 = (0.2, 0.2, 0.0, 0.5, 0.0, 0.1). Suppose further that we have an actual observation vector as d = (0, 0, 0, 1, 0, 0). Then the PSs of p^1 and p^2 are 0.54 and 0.34; while the RPSs of p^1 and p^2 are 0.27 and 0.38, respectively. Even though the PS of p^2 is lower (better), the RPS of p^2 is worse than that of p^1.

Application to Daily Cotton Prices

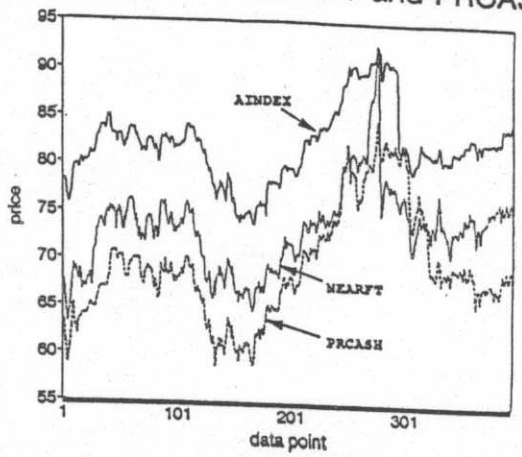
Below we consider turning point forecasts of daily cotton prices. The application is presented in three subsections. First we describe the data. Second we summarize the time series method used to generate probability forecasts. Finally we present our results.

Data

Daily observations on cotton futures prices from the New York Cotton Exchange (NEARFT), daily cotton cash prices from the Memphis market (PRCASH), and daily world cotton prices (AINDEX) are studied. The period of analysis is from June 16, 1989 to December 21, 1990. The data sources are Cotton Price Statistics and Weekly Cotton Market Review published by Agricultural Marketing Service of U.S. Department of Agriculture.

The daily cotton futures prices are daily settlement prices of nearby futures contract in the New York Cotton Exchange for base grade cotton. The Memphis cotton price represents the price of cotton spot market in the United States. It is the price of cotton for grade 41 (Spotted Light Middling) and staple 34 (1-1/16-inch). The world cotton price is "A" Index which is the average of the five cheapest growths of Middling 1-3/32-inch cotton offered for sale in Northern Europe and is considered to be the most representative world cotton price series. Thus, NEARFT represents predictive cotton prices in nearby contract months, and PRCASH represents spot market prices in U.S., while AINDEX represents spot market prices in world. The data are plotted in figure 1.

Figure 1. Plots of the Series AINDEX, NEARFT and PRCASH (data points 1-395).



distributed with mean $\phi(B)$ and variance/covariance VCV = PP'. Here P is a decomposed matrix of VCV. A particular draw of parameter matrix $\phi^*(B)$ is given as:

$$\phi^*(B) = \phi(B) + Pe,$$

where e is a vector of standard normal draws. Uncertainty in residuals is modelled by drawing an error vector ϵ_t^* from the empirical distribution of reshuffled error vectors. Applying the chain rule of forecasting (see Sargent, 1979, p.268), a k-step-ahead forecast vector is generated (observed values of the X vector are substituted for forecasted values where the former are available):

$$X_{t+k}^* = \phi^*(B)X_{t+k-1}^* + \epsilon_{t+k}^*,$$

where X_{t+k}^* is a simulated value for X_{t+k} and ϵ_{t+k}^* is a draw from the one-step error distribution.

Table 2. FPE[a] Statistics on First-Differenced Data.

Lags	Series[b] AINDEX	NEARFT	PRCASH
0	0.3889	0.8220	0.7434
1	0.3861*	0.8098*	0.7278
2	0.3897	0.8185	0.7354
3	0.3907	0.8250	0.7312
4	0.3946	0.8283	0.7368
5	0.3980	0.8140	0.7249*
6	0.4017	0.8190	0.7327

Note: Asterisks denote the smallest FPE statistics for each variable.

[a] $FPE(k) = \frac{T+k+1}{T-k-1} \sum_{t=k+1}^{T} (y_t - \hat{y}_t(k))^2/T$, where T is the number of observations, and k is a particular order of lags used to represent y_t.

[b] The data points vary with respect to lags used. For the lag 0, data points 2-195 were used. For the lag 1 through the lag 6, data points 8-195 were used.

Table 3. Lags on Each Variable in the Autoregressive Representation of Each Variable as Determined Using Hsiao's Search Procedure.

Lags on series	Equation AINDEX	NEARFT	PRCASH
		(lags)	
AINDEX	1	0	0
NEARFT	2	0	0
PRCASH	1	5	0

Note: Data points 7 - 195 were used to calculate multiple FPE statistics.

Prequential Models of Cotton Prices

Plots of the series suggest that they are mean nonstationary (figure 2). Dickey-Fuller and Augmented Dickey-Fuller tests are used to test the null hypothesis that the original series is a random walk (nonstationary). Table 1 shows the t-statistics of the test in levels. From the results it is concluded that these series are nonstationary in levels; while first differenced series appear to be stationary. Models will be built using the first differenced series.

Table 1. t-statistics of Stationarity Tests on Data Points 1 - 195.

	$(1-B)^1$		$(1-B)^2$	
Series	DF[a]	ADF[b]	DF	ADF
AINDEX	-1.192	-2.003	-18.502	-13.857
NEARFT	-1.040	-2.982	-12.757	-14.215
PRCASH	-1.384	-1.865	-16.376	-14.384

[a] Dickey-Fuller (DF) test is on the coefficient ϕ_1, from regression $(1-B)X_t = \phi_0 + \phi_1 X_{t-1}$, where B is the lag operator. Failure to reject the null hypothesis ($\phi_1 = 0$) is consistent with the process X_t being generated as a random walk. As the distribution theory underlying these tests is nonstandard, Monte Carlo generated critical value ($\alpha = 0.05$) of 3.4 will be used (Dickey and Fuller 1979).

[b] The augumented Dickey-Fuller (ADF) test is of the same form as that given in the DF test except lags of the dependent variable as are specified (using FPE) on the right-hand side of each equation.

Previous results have suggested that univariate models perform well (Bessler and Kling, 1986). However, they can not explain the inter-relationships among variables because they have only one variable in the model. But, univariate models may be helpful in providing a good set of forecasts which can be compared to those generated from other, multiple series, models.

A statistical loss function is used for determining the order of lags in an autoregressive model. Table 2 gives estimated FPE statistics on lags 0 through 6 on each series. The minimum FPE-statistics are obtained at lag 1 for AINDEX, at lag 1 for NEARFT, and at lag 0 for PRCASH.

To identify a restricted VAR, the FPE loss function is again applied. Hsiao provided a search procedure to select the lags of which the coefficients are non-zero (Hsiao, 1979). Table 3 gives a summary of lags selected. The maximum lag length was set as five. PRCASH series has no lags (cash prices are generated as a random walk). NEARFT has five lags of PRCASH series only. The AINDEX series has one lag of itself, two lags of NEARFT and one lag of PRCASH in its multivariate representation.

A bootstrap-like procedure was applied to generate predictive probability distribution. Its explicit description is given in Kling and Bessler (1989). Assume a finite AR model: $X_t = \phi(B)X_{t-1} + \epsilon_t$. The procedure followed allows for uncertainty in $\phi(B)$ and ϵ_t. At each date, the elements of $\phi(B)$ are assumed to be normally

Doing this one hundred times generates a set of probability distributions at each horizon, k. The model is then updated at the next data point with the Kalman filter and a new set of probability forecasts is generated with the same procedure.

The prequential turning point prediction has been applied to cotton price series over data points 196 to 295. Several input requirements must be specified in order to make the procedure operational.

i τ is selected as two. It is assumed that we are interested in a directional change in the short-run. As we have first-differenced series, if a sum of forecasts of the series at data point t-1 and t is greater than zero, this implies a forecast of an increase from t-2 to t, and if a sum of forecasts of the series at data point t+1 and t+2 is less than zero, this implies aforecast of a decrease from t to t+2. Thus, t is predicted as a peak. The reverse case is predicted as a trough.

ii Errors are drawn from the empirical distribution of the residuals. Values for the regression coefficients are drawn from a multivariate normal distribution with mean vector ϕ_i and variance Σ_ϕ.

iii 200 is selected as the number of draws N.

iv 20 is selected as the number of periods in the forecast horizon. A twenty period horizon is judged long enough to have at least one turning point.

The univariate model and the restricted VAR model were estimated using data point 1 through 195. One-step-ahead through twenty-step-ahead forecasts were generated at each data point 196 through 295. The predictive distribution W_{it}^P and W_{it}^T and their probability distribution P_{it}^P and P_{it}^T were generated using the above procedure.

To evaluate the forecasting performance, chi-squared statistics are calculated. The calibration functions are plotted to check their underconfidence and overconfidence. The PS, RPS, and SHP are calculated to compare the performances.

Results

The results of UNIV and RVAR forecasts are reported in table 4. If we compare the performance of the univariate model and the restricted VAR model, the restricted VAR model dominates in AINDEX peak and trough and NEARFT trough prediction. As the univariate and mutivariate model on PRCASH are the same, entries in these columns are identical. Seven of twelve chi-squared statistics are greater than 21.6 (critical value at .01 significance level). The hypothesis that forecasts are well-calibrated is rejected for those predictions. The calibration functions are plotted in figure 2 for AINDEX, NEARFT, and PRCASH, respectively. Locations are overestimated for AINDEX peak and trough prediction for UNIV model. This implies that turning points are predicted later than the dates on which turning points actually occur. Locations are underestimated for NEARFT peak prediction. Other locations are close to the actual locations. The functions show flat-steep or flat-steep-flat patterns, which suggests that spread is overestimated, i.e., forecasts are underconfident.

Recalibration was performed on the next data set from data point 296 to 395. To do this we used the miscalibration observed over the period 196 to 295 to adjust issued probabilites over the period 296 to 395.

Table 4. Results of Turning Points Forecasts (Data Points 196 - 295)

Series	Measure	Peak Forecast		Trough Forecast	
		UNIV	RVAR	UNIV	RVAR
AINDEX	PS	0.8046	0.7510*	0.7896	0.6816*
	RPS	1.2062	1.1114*	1.2065	1.0384*
	SHP	0.8435	0.7725*	0.8496	0.7726*
	CHI-SQ	41.71	7.75*	44.51	18.55*
NEARFT	PS	0.8676*	0.8692	0.8300	0.7927*
	RPS	1.7196*	1.7315	1.2143	1.1450*
	SHP	0.8544	0.8146*	0.8540	0.8355*
	CHI-SQ	20.22*	38.57	25.88	22.75*
PRCASH	PS	0.8445	0.8445	0.8192	0.8192
	RPS	1.4601	1.4601	1.2807	1.2807
	SHP	0.8520	0.8520	0.8474	0.8474
	CHI-SQ	26.08	26.08	24.83	24.83

Note: Asterisk denotes the better measure between UNIV model and RVAR model.

The performance results of recalibrated and non-recalibrated forecasts are shown in table 5 and table 6 for turning point peak and turning point trough, respectively. Non-recalibrated models perform better in accuracy for all series except for AINDEX series. Recalibrated UNIV model performs best in accuracy in AINDEX peak prediction. Recalibrated RVAR performs best in accuracy for AINDEX trough prediction. Recalibrated models show the better chi-squared statistics in AINDEX prediction.

In NEARFT prediction, non-recalibrated UNIV model shows the most accurate forecasts for peak and trough prediction. Chi-squared statistics show that they are all well-calibrated except for recalibrated RVAR trough prediction. In PRCASH prediction, non-recalibrated UNIV model performs best in accuracy for peak prediction while non-recalibrated RVAR model performs best in accuracy for trough prediction.

All forecasts for PRCASH turning points fail in chi-squared tests. Recalibration has improved chi-squared statistics in AINDEX series only. Non-recalibrated models fail in chi-squared tests for AINDEX prediction. Eight out of twelve recalibrated forecasts fail in chi-squared tests and eight out of twelve non-recalibrated forecasts fail in chi-squared tests at .01 significance level.

Summary and Discussion

Prequential analysis was applied to turning point forecasts for three cotton price series: world market price (AINDEX), the nearby futures price (NEARFT), and Memphis cash market price (PRCASH). As "prequential" implies, sequential, ex ante, probability forecasting was performed.

Figure 2. Calibration Functions of AINDEX, NEARFT, PRCASH Turning Point Prediction (data points 196-295).

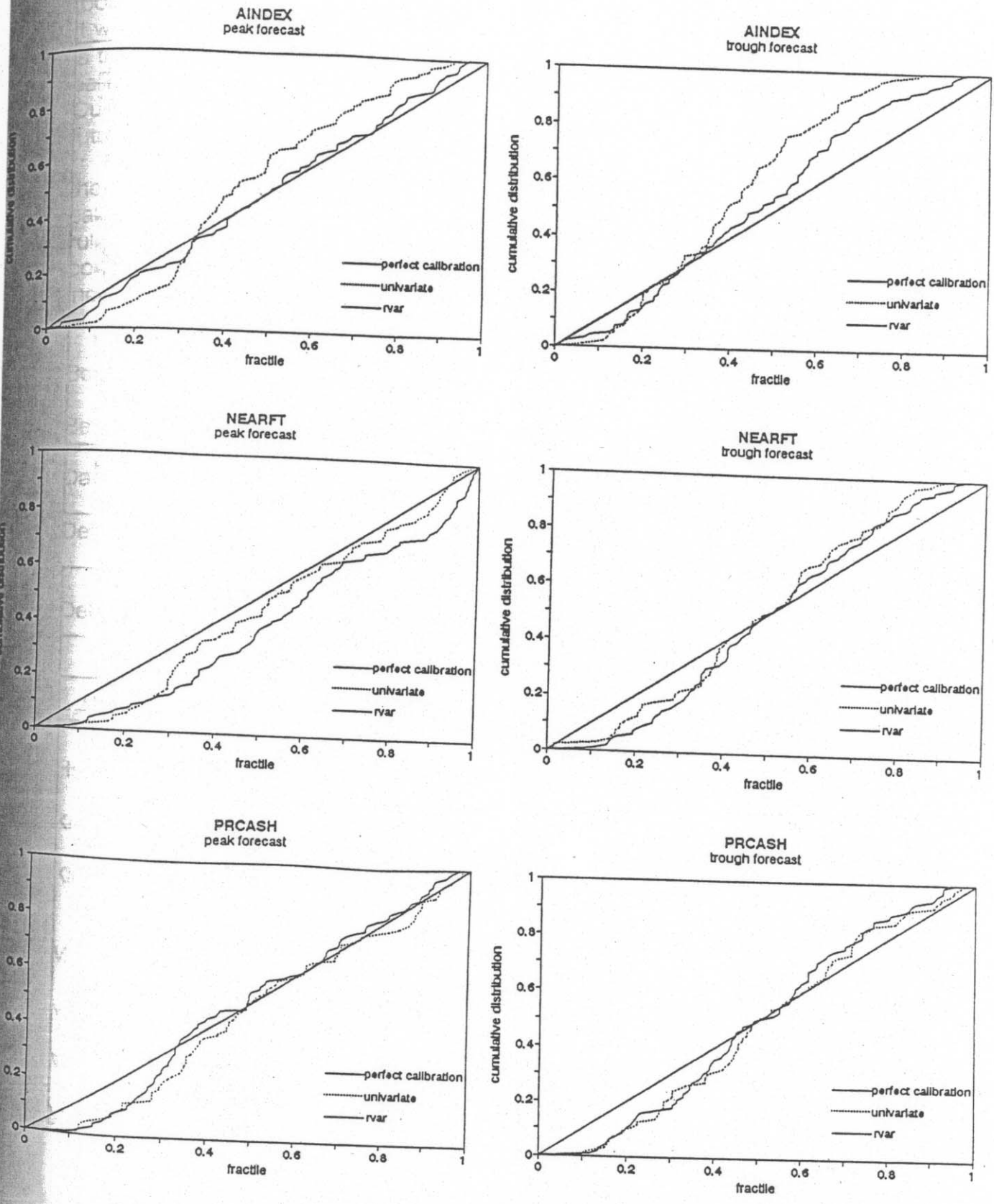


Table 5. Results of Turning Point Peaks Forecasts: Recalibrated Vs. Non-recalibrated (Data Points 296 - 395)

		UNIV		RVAR	
Series	Measure	RECAL	NONRECAL	RECAL	NONRECAL
INDEX	PS	0.8045	0.7503*	0.7540	0.7354*
	RPS	1.2092*	1.3882	1.2294*	1.2418
	SHP	0.7877*	0.8500	0.7925*	0.8053
	CHI-SQ	29.91*	59.39	31.97*	33.82
NEARFT	PS	0.8722	0.8542*	0.8911	0.8594*
	RPS	1.6977	1.6470*	1.7480	1.6802*
	SHP	0.8615	0.8585*	0.8476	0.8218*
	CHI-SQ	17.33	16.92*	13.17*	15.04
PRCASH	PS	0.8411	0.8238*	0.8411	0.8238*
	RPS	1.2336*	1.2449	1.2336*	1.2449
	SHP	0.8323*	0.8507	0.8323*	0.8570
	CHI-SQ	32.96	32.33*	32.96	32.33*

Note: Asterisk denotes the best measure among recalibrated and non-recalibrated UNIV model and RVAR model.

Table 6. Results of Turning Point Troughs Forecasts: Recalibrated Vs. Non-recalibrated (Data Points 296 - 395).

		UNIV		RVAR	
Series	Measure	RECAL	NONRECAL	RECAL	NONRECAL
AINDEX	PS	0.8022	0.7850*	0.7185	0.6940*
	RPS	0.9399	0.9399*	1.1881	0.8459*
	SHP	0.9399*	1.1881	0.8459*	0.9362
NEARFT	PS	0.8875	0.8525*	0.9616	0.9028*
	RPS	1.5213	1.4749*	1.6118	1.5128*
	SHP	0.8173*	0.8471	0.7562*	0.8056
	CHI-SQ	16.08	14.24*	35.67	17.92*
PRCASH	PS	0.9278	0.8918*	0.9278	0.8918*
	RPS	2.4489	2.3093*	2.4489	2.3093*
	SHP	0.8152*	0.8362	0.8152*	0.8362
	CHI-SQ	29.14*	30.78	29.14*	30.78

Note: Asterisk denotes the best measure among recalibrated and non-recalibrated UNIV model and RVAR model.

As traditional point forecast evaluation tools such as mean-squared error (MSE) can not be used without ignoring much potentially useful information, other evaluation tools were considered. They include PS, RPS, sharpness, and chi-squared statistics. It was argued that the RPS is a more appropriate measure for "ordered" variables than is the PS. Chi-squared goodness-of-fit statistics are used to formally test the calibration of forecasts. However, the chi-squared statistics do not capture accuracy. Our goal of probability forecasting is to issue accurate and sharp distributions for the future events at each data point. This is measured by the RPS.

It is proposed that the chi-squared test be used to test calibration formally, and the scoring rules be used to select the better performing models. That is to say, calibration is not the only criterion of probability assessment "goodness." Scoring rules offer us more information and should be used. Ultimately, forecast users must consider the rate at which they will trade off reliable probabilites (good calibration) for increased precision (sharper forecasts).

References

Bessler, D. A., and J. L. Kling. "Forecasting Vector Autoregression with Bayesian Priors." *American Journal of Agricultural Economics* 68(1986):144-51.

Bessler, D. A., and J. L. Kling. "Prequential Analysis of Cattle Prices." *Applied Statistics* 39(1):95-106.

Dawid, A. P. "Statistical Theory: The Prequential Approach." *Journal of Royal Statistical Society* 147(1984):278-92.

DeGroot, M. H., and S. E. Fienberg. "Assessing Probability Assessors: Calibration and Refinement." *Statistical Decision Theory and Related Topics III*, Vol. 1, eds. S.S. Gupta and J. O. Berger, pp. 291-314. New York: Academic Press, Inc., 1982.

DeGroot, M. H. and S. E. Fienberg. "Comparing Probability Forecasters: Basic Binary Concepts and Multivariate Extensions." *Bayesian Inference and Decision Techniques*, eds. P. Goel and A. Zellner, New York: Elsevier Science Publisher, 1986.

Epstein, E. S. "A Scoring System for Probability Forecasts of Ranked Categories." *Journal of Applied Meteorology* 8(1969).

Hsiao, C. "Autoregressive Modeling of Canadian Money and Income Data." *Journal of the American Statistical Association* 74(1979):553-60.

Kling, J. L. "Predicting the Turning Points of Business and Economic Time Series." *Journal of Business* 60(1987):201-38.

Kling, J. L., and D. A. Bessler. "Calibration-based Predictive Distributions: An Application of Prequential Analysis to Interest Rates, Money, Prices, and Output." *Journal of Business* 62(1989):477-500.

Murphy, A. H. "The Ranked Probability Score and the Probability Score: A Comparison." *Monthly Weather Review* 98(1970):917-24.

Murphy, A. H. "Scalar and Vector Partitions of the Probability Score: Part I. Two-State Situation." *Journal of Applied Meteorology* 11(1972a):273-282.

Murphy, A. H. "Scalar and Vector Partitions of the Probability Score: Part II. N-State Situation." *Journal of Applied Meteorology* 11(1972b):1183-192.

Murphy, A. H. "Scalar and Vector Partitions of the Ranked Probability Score." *Monthly Weather Review* 100(1972c):701-08.

Murphy, A. H. "A New Vector Partition of the Probability Score." Journal of Applied Meteorology 12(1973):595-600.

Neftci, S. N. "Optimal Prediction of Cyclical Downturns." *Journal of Economic Dynamics and Control* 4(1982):225-41.

Sargent, T. *Macroeconomic Theory*. New York: Academic Press, 1979.

Wecker, W. F. "Predicting the Turning Points of a Time Series." *Journal of Business* 52(1979):35-50.

Winkler, R. L. "On Good Probability Appraisers." *Bayesian Inference and Decision Techniques*, eds., P. Goel and A. Zellner, New York: Elsevier Science Publisher, 1986.

Zarnowitz, V., and G. H. Moore. "Sequential Signals of Recession and Recovery." *Journal of Business* 55(1982):57-85.

Zellner, A., C. Hong, and C. Min. "Forecasting Turning Points in International Output Growth Rates Using Bayesian Exponentially Weighted Autoregression, Time-Varying Parameter, and Pooling Techniques." *Journal of Econometrics* 49(1991):275-304.