UNBIASEDNESS OF PREDICTIONS FROM ESTIMATED AUTOREGRESSIONS WHEN THE TRUE ORDER IS UNKNOWN[1]

BY JEAN-MARIE DUFOUR

1. INTRODUCTION

AUTOREGRESSIVE MODELS have become more and more widely used for the purpose of making forecasts in econometrics and other fields. A question which comes out naturally here is whether the forecasts obtained in this way are "unbiased." By an unbiased forecast, we mean a forecast such that the prediction error has mean zero or, if the mean does not exist, has median zero.

It is well known that minimum mean-square-error predictors yield unbiased forecasts, at least when their functional form is either unconstrained or linear (see Kendall and Stuart [20, Chapters 26–28], Box and Jenkins [9, Chapter 5]). Derivation of such optimal predictors usually requires exact knowledge of the joint distribution of the variables involved, or at least some of its parameters. However, when parameters must be estimated, it is not clear that the same property holds, or any of the standard properties of optimal forecasts.[2] In the important case of the classical linear model (with strictly exogenous regressors), unbiasedness can indeed be proved using the unbiasedness of least squares estimates and the exogeneity of the regressors (see Theil [27, pp. 122–124]). But in more complex situations where unbiased coefficient estimates are not usually available (e.g. models with lagged dependent variables), such simple proofs seem precluded.

In this paper, we will consider the case of estimating the parameters of an autoregressive model by ordinary least squares. It is well known that the coefficient estimates obtained by this method are generally biased (see Hurwicz [19], White [28], Shenton and Johnson [26], Orcutt and Winokur [22], Sawa [24], Bhansali [7]). This leads one to expect that the forecasts obtained using such estimates will also be biased. But this is not always the case: it can be shown that, for an autoregressive process of order one,

$$(1.1) \qquad x_t = \alpha_0 + \alpha_1 x_{t-1} + \epsilon_t \qquad\qquad (t = 1, 2, \ldots),$$

$$\qquad\qquad = x_0 \qquad\qquad\qquad\qquad (t = 0),$$

the forecast errors obtained by estimating (1.1) with ordinary least squares have zero mean (when the mean exists), provided the innovations ϵ_t are independent and identically distributed random variables with zero mean, finite variance, and symmetric distribution, and x_0 is a symmetrically distributed random variable independent of ϵ_t, $t = 1, 2, \ldots$, with finite variance and some appropriate mean μ. Specifically, Malinvaud [21, p. 582] has stated this result for the case where $\alpha_0 = \mu = 0$, when no constant is included in the fitted model; on the other hand, Fuller and Hasza [17] have proved it for the following cases (with a constant estimated): (i) $|\alpha_1| < 1$ and $\mu = \alpha_0/(1 - \alpha_1)$; (ii) $\alpha_0 = \mu = 0$; (iii) $\alpha_0 = 0$ and $\alpha_1 = 1$ (with no other condition on x_0, except independence from ϵ_t, $t = 1, 2, \ldots$). The results of these authors were obtained under the assumption that the "true" model and the estimated model are the same.

[2] For a discussion illustrating the importance of this question in the context of "rational expectations" macroeconomic models, see Friedman [16].

It is natural to ask whether the forecasts of an autoregressive model are unbiased when the true and estimated models are different, notably when the model estimated is too "parsimonious." The main purpose of this paper is to give a rather weak condition, namely a condition of "joint symmetry" of the distribution of the process (to be defined more precisely below), under which an autoregressive model, of arbitrary order, estimated by ordinary least squares, will yield unbiased forecasts. An important implication of this result is that, under quite general circumstances, forecasts will be unbiased even if the order of the autoregressive model estimated is *lower* than the actual one. Further, the same property will hold a fortiori when the order is correct or too high.

We may note also that the results presented below hold exactly, even in finite samples. Previous studies of the properties of prediction errors from estimated time-series models considered large-sample properties and focused on the mean squared error of prediction. Work on these issues includes Davisson [13], Akaike [1], Box and Jenkins [9, Section A7.3], Cleveland [10], Bloomfield [8], Bhansali [5, 6, 7], Schmidt [25], Yamamoto [30, 31], Baillie [2, 3, 4], Phillips [23], Fuller and Hasza [17, 18], Davies and Newbold [12], and Yajima [29]. Of course, the fact that an autoregressive model produces unbiased forecasts does not imply it is "optimal," in the sense of producing forecasts with minimum mean-square error. However, the property of unbiasedness is a basic property of any forecasting procedure and it is important to know that it will hold exactly under surprisingly weak conditions and despite frequently encountered specification errors. On the other hand, the mean-square prediction error will clearly be affected if the model fitted contains either not enough or too many parameters. Incidentally, in finite samples, the best strategy is not always to estimate the "right" model but may involve estimating a more parsimonious model (see Bhansali [7, Section 6]). This is due to the possibility that large noise is introduced by parameter estimation. Our result thus implies that, under rather wide circumstances, choice of a more parsimonious model will not introduce any bias in the forecasts. For further details on the impact of estimating a misspecified time-series model on the mean-square prediction error, the reader may consult Davisson [13], Cleveland [10], Bloomfield [8], Cogger [11], and Bhansali [7].

The basic definitions used in the paper are given and explicated in Section 2, while the theorem is stated and proved in Section 3. The main conclusions are summarized in Section 4.

2. FRAMEWORK

The condition we shall use in order to show that unbiasedness holds is that of *joint symmetry* of a stochastic process about some constant. It is defined as follows.

DEFINITION: A stochastic process $\{x_t : t \in I\}$, where I is a subset of the integers and x_t is real, is said to be jointly symmetric about a given constant μ if and only if the processes $\{x_t - \mu : t \in I\}$ and $\{\mu - x_t : t \in I\}$ have the same joint distribution.

It is very easy to see that, if $\{x_t : t \in I\}$ is jointly symmetric about μ, then $\{x_t : t \in I_1\}$ is also jointly symmetric about μ for any subset $I_1 \subseteq I$. Further, when $\{x_t : t \in I\}$ contains only one random variable, say x_1, joint symmetry is equivalent to stating that x_1 has a distribution symmetric about μ. It is important to note that this condition is quite distinct from that of "exchangeability" of random variables in the process (see Feller [15, p. 225]). The random variables are ordered by the index set I and cannot generally be permuted without the joint distribution being modified (unless special conditions hold); furthermore, the marginal distributions of the x_t's are not necessarily identical.

A simple case where the joint symmetry property holds is the one where $\{x_t : -\infty < t < +\infty\}$ can be expressed as a moving average of symmetrically distributed innovations,

i.e.,

$$x_t = \mu + \sum_{i=0}^{\infty} \Theta_i \epsilon_{t-i} \qquad\qquad (t = \ldots, -1, 0, 1, \ldots),$$

where the variables ϵ_t are independent with distributions symmetric about zero. Clearly, all stationary processes with independent and symmetrically distributed innovations (e.g., Gaussian stationary processes) are jointly symmetric. Similarly, if x_t, $t = 1, 2, \ldots$, can be expressed as an autoregression of order p (possibly infinite):

$$x_t = \mu + \sum_{i=1}^{p} \alpha_i (x_{t-i} - \mu) + \epsilon_t \qquad\qquad (t = 1, 2, \ldots),$$

where $\epsilon_1, \epsilon_2, \ldots$, are independent with distributions symmetric about zero and $(x_{-p+1}, \ldots, x_0)$ is jointly symmetric about μ as well as independent of ϵ_t, $t = 1, 2, \ldots$, then $\{x_t : t = 1, 2, \ldots\}$ is jointly symmetric about μ. Further, we may note it is not necessary to assume that the roots of the equation $1 - \sum_{i=1}^{p} \alpha_i z^i = 0$ be outside the unit circle. Thus, we are not limiting ourselves to covariance-stationary processes.

We shall now consider the problem of predicting a process symmetric about an arbitrary constant μ using an autoregressive model of order $p \geq 1$ estimated with an intercept term.

3. THEOREM

Let $\{x_t\}$ be a process symmetric about some unknown constant μ. Observations on x_t, $t = -p + 1, \ldots, 0, \ldots, T$, are available and we wish to predict x_{T+l}, $l = 1, 2, \ldots$, using an autoregressive model of order p:

$$(3.1) \qquad x_t = \alpha_0 + \sum_{k=1}^{p} \alpha_k x_{t-k} + u_t \qquad\qquad (t = 1, \ldots, T).$$

If the coefficient vector $\beta = (\alpha_0, \alpha')'$, where $\alpha = (\alpha_1, \ldots, \alpha_p)'$, is estimated by ordinary least squares, the estimate of β obtained is given by

$$(3.2) \qquad \hat{\beta}_T = \left(\bar{X}_T' \bar{X}_T \right)^{-1} \bar{X}_T' x_T^T,$$

where

$$\bar{X}_T = [i_T, X_T],$$

$$(3.3) \qquad x_t^T = (x_{t-T+1}, x_{t-T+2}, \ldots, x_t)', \qquad t \geq T - p,$$

$$X_T = \left[x_{T-1}^T, x_{T-2}^T, \ldots, x_{T-p}^T \right],$$

and $i_T = (1, 1, \ldots, 1)'$ is the $T \times 1$ unit vector. In what follows, we will assume that $\hat{\beta}_T$ exists with probability 1. Then the standard least squares forecast of x_{T+l} is

$$(3.4) \qquad \hat{x}_T(l) = \hat{\alpha}_{0,T} + \sum_{k=1}^{p} \hat{\alpha}_{k,T} \hat{x}_T(l - k), \qquad l \geq 1,$$

where $\hat{\alpha}_{k,T}$ $(k = 0, 1, \ldots, p)$ is the kth component of $\hat{\beta}_T$ and

$$(3.5) \qquad \hat{x}_T(l) = x_{T+l}, \qquad \text{if } l \leq 0.$$

We will now show that the errors of prediction $x_{T+l} - \hat{x}_T(l)$ have distributions symmetric about zero and thus have mean zero (provided this mean exists).

THEOREM: *Let the stochastic process* $\{x_t : -p + 1 \leq t \leq T + L\}$ *be jointly symmetric about a certain constant* μ $(p \geq 1, T \geq 2, L \geq 1)$. *Then, if the model* (3.1) *is estimated by ordinary least squares, each error of prediction* $e_T(l) \equiv x_{T+l} - \hat{x}_T(l)$, $1 \leq l \leq L$, *where* $\hat{x}_T(l)$ *is given by* (3.2) *and* (3.4), *has a distribution symmetric about zero. Furthermore, if* x_{T+l} *and* $\hat{x}_T(l)$ *have finite means, then*

$$E[x_{T+l} - \hat{x}_T(l)] = 0 \qquad\qquad (l = 1, \ldots, L).$$

PROOF: Let $\eta = (x_{-p+1}, \ldots, x_{T+L})'$, $M = T + L + p$, and $i_M = (1, 1, \ldots, 1)'$ the $M \times 1$ unit vector. By the joint symmetry (about μ) condition, the vectors $\eta - \mu i_M$ and $\mu i_M - \eta$, have the same distribution. Further, by partitioning $\hat{\beta}_T = (\hat{\alpha}_{0,T}, \hat{\alpha}_T')'$, we get

$$\hat{\alpha}_T = (X_T' A_T X_T)^{-1} X_T' A_T x_T^T,$$

$$\hat{\alpha}_{0,T} = \frac{1}{T} \sum_{t=1}^{T} \left[x_t - \sum_{k=1}^{p} \hat{\alpha}_{k,T} x_{t-k} \right],$$

where $\hat{\alpha}_T = (\hat{\alpha}_{1,T}, \ldots, \hat{\alpha}_{k,T})'$ and $A_T = I_T - (1/T) i_T i_T'$. Defining

$$\tilde{X}_T = X_T - \mu i_T i_p', \qquad \tilde{x}_T^T = x_T^T - \mu i_T,$$

we have

$$A_T \tilde{X}_T = A_T X_T, \qquad A_T \tilde{x}_T^T = A_T x_T^T;$$

hence, using the fact that A_T is idempotent,

$$\hat{\alpha}_T = (\tilde{X}_T' A_T \tilde{X}_T)^{-1} \tilde{X}_T' A_T \tilde{x}_T^T \equiv \hat{a}_T(\eta - \mu i_M),$$

from which it is easy to see that $\hat{a}_T(\cdot)$ is an even function of $\eta - \mu i_M$:

(3.6) $\hat{a}_T(\eta - \mu i_M) = \hat{a}_T(\mu i_M - \eta)$.

On the other hand,

$$\hat{\alpha}_{0,T} = \tilde{\alpha}_{0,T} + \mu \left[1 - \sum_{k=1}^{p} \hat{\alpha}_{k,T} \right],$$

where

$$\tilde{\alpha}_{0,T} = \frac{1}{T} \sum_{t=1}^{T} \left[(x_t - \mu) - \sum_{k=1}^{p} (x_{t-k} - \mu) \hat{\alpha}_{k,T} \right]$$

$$\equiv \tilde{a}_{0,T}(\eta - \mu i_M).$$

We can see easily that $\tilde{a}_{0,T}(\cdot)$ is an odd function of $\eta - \mu i_M$:

(3.7) $\tilde{a}_{0,T}(\eta - \mu i_M) = -\tilde{a}_{0,T}(\mu i_M - \eta)$.

Then, we can rewrite the forecasts (3.4) in the form:

$$\hat{x}_T(l) - \mu = \tilde{\alpha}_{0,T} + \sum_{k=1}^{p} [\hat{x}_T(l-k) - \mu]\hat{\alpha}_{k,T}$$

$$\equiv F_l(\eta - \mu i_M) \qquad\qquad (l = 1, \ldots, L).$$

For $l = 1$, it is clear, using (3.6) and (3.7), that $F_l(\cdot)$ is an odd function of $\eta - \mu i_M$; hence, by induction,

$$F_l(\eta - \mu i_M) = -F_l(\mu i_M - \eta) \qquad\qquad (l = 1, \ldots, L).$$

The forecast errors $e_T(l) = x_{T+l} - \hat{x}_T(l)$ can now be expressed as

$$e_T(l) = (x_{T+l} - \mu) - F_l(\eta - \mu i_M) \equiv G_l(\eta - \mu i_M);$$

hence,

$$G_l(\eta - \mu i_M) = -G_l(\mu i_M - \eta).$$

Thus, from the symmetry condition, $e_T(l)$ and $-e_T(l)$ must have the same distribution, i.e. $e_T(l)$ has a distribution symmetric about zero. Furthermore, when the means of x_{T+l} and $\hat{x}_T(l)$ exist, this implies

$$E[x_{T+l} - \hat{x}_T(l)] = 0 \qquad\qquad (l = 1, \ldots, L).$$

$$Q.E.D$$

An important implication of the above theorem is that, for jointly symmetric processes, any autoregressive model which includes an intercept term will yield unbiased forecasts, when the coefficients of the autoregression are estimated by ordinary least squares. This is stated in the following corollary.

COROLLARY: *Let* $\{x_t\}$ *be a stochastic process having a representation (at least for $t \geq 1$) of the form*:

$$x_t = \beta_0 + \sum_{k=1}^{p^*} \beta_k x_{t-k} + \epsilon_t \qquad\qquad (t = 1, 2, \ldots)$$

where $p^* \geq 1$. Assume that $\sum_{k=1}^{p^*} \beta_k \neq 1$ whenever $\beta_0 \neq 0$, let μ_0 be a real constant, and

$$\mu = \beta_0 \Big/ \left(1 - \sum_{k=1}^{p^*} \beta_k\right), \qquad if \quad \sum_{k=1}^{p^*} \beta_k \neq 1,$$

$$= \mu_0, \qquad\qquad if \quad \sum_{k=1}^{p^*} \beta_k = 1 \quad and \quad \beta_0 = 0.$$

Let the random variables ϵ_t be independent with distributions symmetric about zero. Let also p be a positive integer (possibly different from p^*), $\bar{p} = \max\{p, p^*\}$ and $(x_{-\bar{p}+1}, \ldots, x_0)$ be jointly symmetric about μ as well as independent of $\{\epsilon_t : t = 1, 2, \ldots\}$. Then, if an autoregressive model of order p, as given in (3.1), is fitted to x_t ($t = 1, \ldots, T$) by ordinary least squares, the errors of prediction $x_{T+l} - \hat{x}_T(l)$, where $\hat{x}_T(l)$ is given by (3.4) and $l \geq 1$,

have distributions symmetric about zero. Furthermore, if x_{T+l} and $\hat{x}_T(l)$ have finite means,

$$E[x_{T+l} - \hat{x}_T(l)] = 0 \qquad\qquad (l = 1, 2, \dots).$$

PROOF: The result follows from the Theorem and the observation that the stochastic process $\{x_t : t = -\bar{p} + 1, -\bar{p} + 2, \dots\}$ is jointly symmetric about μ.   *Q.E.D.*

The above theorem holds for all values of μ provided the autoregression estimated contains an intercept term. When $\mu = 0$, similar results can be shown to hold when one estimates an autoregressive model with no intercept (for a proof, see Dufour [14]).

4. SUMMARY

In this paper a wide class of stochastic processes enjoying a simple symmetry property was first defined: we called these "jointly symmetric processes." They include as special cases: Gaussian stationary processes, processes representable as moving averages of independent symmetrically distributed innovations and a large class of autoregressive processes, stationary as well as nonstationary. Then, we showed that any autoregressive model estimated by ordinary least squares yields unbiased forecasts when applied to a jointly symmetric process. Specifically, for a process symmetric about an arbitrary constant μ, any autoregressive model of order $p \geq 1$ which includes an intercept will yield unbiased forecasts: prediction errors have distributions symmetric about zero and thus have zero mean (whenever the mean exists), when the model is estimated by ordinary least squares. The proof of this property is based on showing that each prediction error is an odd function of the centered process $x_t - \mu$. A similar approach was used previously by Fuller and Hasza [17], on predictors for the first-order autoregressive process when the model estimated is correct; we show here that the unbiasedness property holds under much weaker conditions and for a larger class of predictors. An important implication of the theorem is that, for a symmetric autoregressive process, one will get unbiased forecasts *even if the order of the estimated model is lower than the actual one.* These conclusions hold exactly in finite samples. Furthermore, they do not depend on the usual normality assumption.

Université de Montréal

REFERENCES

[1] AKAIKE, H.: "Statistical Predictor Identification," *Annals of the Institute of Statistical Mathematics*, 22(1970), 203–217.

[2] BAILLIE, R. T.: "The Asymptotic Mean Squared Error of Multistep Prediction from the Regression Model with Autoregressive Errors," *Journal of the American Statistical Association*, 74(1979), 175–184.

[3] ———: "Prediction from ARMAX Models," *Journal of Econometrics*, 12(1980), 365–374.

[4] ———: "Prediction from the Dynamic Simultaneous Equation Model with Vector Autoregressive Errors," *Econometrica*, 49(1981), 1331–1337.

[5] BHANSALI, R. J.: "Asymptotic Mean-square Error of Predicting More than One-step ahead Using the Regression Method," *Applied Statistics*, 23(1974), 35–42.

[6] ———: "Linear Prediction by Autoregressive Model Fitting in the Time Domain," *The Annals of Statistics*, 6(1978), 224–231.

[7] ———: "Effects of Not Knowing the Order of an Autoregressive Process on the Mean Squared Error of Prediction-I," *Journal of the American Statistical Association*, 76(1981), 588–597.

[8] BLOOMFIELD, P.: "On the Error of Prediction of a Time Series," *Biometrika*, 59(1972), 501–507.

[9] Box, G. E. P., AND G. JENKINS: *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day, 1970.

[10] CLEVELAND, W. S.: "Fitting Time Series Models for Prediction," *Technometrics*, 13(1971), 713–723.

[11] COGGER, K. O.: "Specification Analysis," *Journal of the American Statistical Association*, 68(1973), 899–905.

[12] DAVIES, N., AND P. NEWBOLD: "Forecasting with Misspecified Models," *Applied Statistics*, 29(1980), 87–92.

[13] DAVISSON, I. D.: "The Prediction Error of Stationary Gaussian Time Series of Unknown Covariance," *IEEE Transactions on Information Theory*, IT-11 (1965), 527–532.

[14] DUFOUR, J.-M.: "A Specification Error Theorem for Predictions from Estimated Autoregressions," Cahier 8132, Centre de recherche en développement économique et Département de science économique, Université de Montréal, 1981.

[15] FELLER, W.: *An Introduction to Probability Theory and its Applications, Volume II*. New York: John Wiley and Sons, 1966.

[16] FRIEDMAN, B. M.: "Optimal Expectations and the Extreme Information Assumptions of 'Rational Expectations' Macromodels," *Journal of Monetary Economics*, 5(1979), 23–41.

[17] FULLER, W., AND D. P. HASZA: "Predictors for the First-order Autoregressive Process," *Journal of Econometrics*, 13(1980), 139–157.

[18] ————: "Properties of Predictors for Autoregressive Time Series," *Journal of the American Statistical Association*, 76(1981), 155–161.

[19] HURWICZ, L.: "Least Squares Bias in Time Series," in *Statistical Inference in Dynamic Economic Models*, ed. by T. C. Koopmans. New York: John Wiley and Sons, 1950, pp. 365–383.

[20] KENDALL, M. G., AND A. STUART: *The Advanced Theory of Statistics, Volume 2*. London: Charles Griffin and Company, 1967.

[21] MALINVAUD, E.: *Méthodes statistiques de l'économétrie*, Second Edition. Paris: Dunod, 1969.

[22] ORCUTT, G. H., AND H. S. WINOKUR: "First Order Autoregression: Inference, Estimation and Prediction," *Econometrica*, 37(1969), 1–14.

[23] PHILLIPS, P. C. B.: "The Sampling Distribution of Forecasts from a First-Order Autoregression," *Journal of Econometrics*, 9(1979), 241–261.

[24] SAWA, T.: "The Exact Moments of the Least Squares Estimator for the Autoregressive Model," *Journal of Econometrics*, 8(1978), 159–172.

[25] SCHMIDT, P.: "The Asymptotic Distribution of Forecasts in the Dynamic Simulation of an Econometric Model," *Econometrica*, 42(1974), 303–309.

[26] SHENTON, R. L., AND W. L. JOHNSON: "Moments of a Serial Correlation Coefficient," *Journal of the Royal Statistical Society Series B*, 27(1965), 308–320.

[27] THEIL, H.: *Principles of Econometrics*. New York: John Wiley and Sons, 1971.

[28] WHITE, J. S.: "Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient," *Biometrika*, 48(1961), 85–94.

[29] YAJIMA, Y.: "On Prediction of Integrated Moving Average Processes," *Annals of the Institute of Statistical Mathematics*, 32(1980), 81–94.

[30] YAMAMOTO, T.: "Asymptotic Mean Square Prediction Error for an Autoregressive Model with Estimated Coefficients," *Applied Statistics*, 25(1976), 123–127.

[31] ————: "Predictions of Multivariate Autoregressive-Moving Average Models," *Biometrika*, 68(1981), 485–492.