WORKING PAPER No. 19
December 1992

PREDICTION WITH A LINEAR REGRESSION MODEL AND ERRORS IN A REGRESSOR

Bo Jonsson

Svensk resumé på sid 27.

Uppsala University
Department of Statistics
P.O. Box 513
S-751 20 Uppsala
Sweden

December 1992

PREDICTION WITH A LINEAR REGRESSION MODEL

AND

ERRORS IN A REGRESSOR

by

Bo Jonsson

This research reported has been supported by the Swedish Council for Research in the Humanities and Social Sciences (HSFR) and by the National Institute of Econimic Research (KI).
Helpful comments from Anders Ågren, Anders Christoffersson, Tor Jacobsson, Wladislaw Milo and Lars-Erik Öller are gratefully acknowledged.

ABSTRACT

The subject under study is prediction with a simple linear regression model in the presence of errors in variables. The paper focuses on the case of a non-stochastic true regressor (x), but stochastic x is also treated. For a wide range of true x values around the mean of x in the estimation period, predictions based on OLS on the observed variables is to be preferred in terms of MSE to a predictor based on consistent estimation of the parameters. This can be so also when x follows a trend and predictions are made for the next observation. When the error variance of the regressor in the prediction period differs from the mean error variance in the estimation period sample, a predictor based on a modified OLS estimator, adjusted for that difference, behaves as the OLS predictor in the case of equal error variances. Under certain assumptions the modified OLS predictor produces consistent predictions, conditional on observed X for the case of stochastic x.

1.INTRODUCTION

Observations on economic time series are often subject to errors; the last data points can be preliminary, the series can contain sampling errors, etc. For a discussion of sources of errors, see for instance Pierce (1981). One example of the case of sampling errors is when variables are based on household survey data concerning attitudes about the economic outlook and buying plans. Such variables are then used as explanatory variables in consumption or investment functions. In regression analysis these errors are often neglected, although it is well known that OLS produces inconsistent estimates of the structural relationship if an independent variable suffers from measurement error. For instance in a situation of a regressor based on survey estimated proportions Jonsson (1992) showed that the large sample bias may be large. Much attention has been paid to methods for estimating the structural parameters in the case of error in variables (for an early review see Madansky (1959) and for recent extensive surveys see e.g. Fuller (1987) and Aigner et al. (1986)). Although prediction is an important task of regression analysis this topic in connection with errors in variables is just marginally touched upon in the literature. Perhaps this is due to a misunderstanding. Fuller (1987, p. 74) expresses this in the following way: "One often hears the statement , 'If the objective is prediction, it is not necessary to adjust for measurement error.' As with all broad statements, this requires a considerable number of conditions to be correct". The subject under study in this paper is that of measurement errors and prediction.

Assume the following model:

Structural equation: $y_t = \alpha + \beta x_t + q_t$, where q_t is $IN(0,\sigma^2_q)$, $t=\{1,..,T,p\}$. (1)

Measurement equations: $Y_t = y_t + w_t$

$X_t = x_t + u_t$.

w_t is $IN(0,\sigma^2_w)$ and u_t is $IN(0,\sigma^2_{ut})$. w_t, q_t and u_t are assumed to be independent. T is the size of the estimation period sample. p denotes the prediction period. If the error variance in X_t is constant, then $\sigma^2_{ut} = \sigma^2_u$.

In one case there seems to be no difficulty in handling the prediction situation. This occurs when X_p is a random selection from the same population that generated the estimation period sample (see e.g. Fuller, 1987, p. 75). OLS will then provide the best linear predictor given X_p, if the error variance in X_t is the same for all t. Further, if x_t is assumed to be $IN(\mu_x,\sigma^2_x)$, (Y_t,X_t) will be distributed as a bivariate normal random variable. Then the OLS predictor is the best unbiased predictor conditional on the observed X. Fuller emphasises that the use of OLS for prediction purposes requires the assumption that X_p be a random selection from the same distribution that generated the X_t of the estimation period sample. An example when this is not the case is given in Ganse et al. (1983). They assume different distributions of the true x in the estimation and prediction period samples, while the other parameters in their model are assumed to be the same for both samples. Another example concerning yield forecasting is found in Reiser et al. (1992) . They argue for using a consistently estimated prediction equation because of possible differences between the populations behind the estimation and prediction period samples.

Wold (1963) has shown that if $Y = \alpha^* + \beta^* X + \varepsilon$ with $E(Y|X) = \alpha^* + \beta^* X$ and if the first and second order sample moments have the corresponding theoretical moments of the variables Y and X as stochastic limits, then the least square regression of Y on X will provide consistent estimates of the parameters α^* and β^*. If x_t and u_t in (1) are normally distributed with constant variances, then $E(Y_t|x_t)$ and $E(Y_t|X_t)$ are both linear in x_t and X_t, respectively. Wold's theorem then implies that the regression of Y_t on X_t will give consistent predictions. Wold especially points out that the assumptions about the sample moments are satisfied in the case of independent replications of a controlled experiment, and in the case when the variables Y and X are given as time series that are stationary and ergodic. Hodges and Moore (1972) also point out the necessity that the values used in making forecasts are drawn from the same stationary distribution and subject to the same sort of error to avoid prediction bias. Unfortunately, as they write, this is often not the case and it is never true when an independent variable follows a trend over time. They find it disturbing that if X_t is unbiased then any bias in the estimate of β "will be transmitted into a biased prediction". They then argue for trying to ensure that X_p and the estimate of β used in the prediction are both unbiased.

Non stationarity can also appear as non-constant error variance in X_t. Given only preliminary data the error variance of the prediction period is greater than that of the estimation period. Another reason would be unequal sample sizes for survey based X_t values. An extreme case occurs when the error variance in the prediction period is zero, what will mean that predictions are made for a value of the true x_p. In that situation OLS is not suitable for prediction, at least not if T is large (see e.g. Malinvaud, 1970, p. 382).

Yum and Neuhardt (1984) examined the prediction problem for model (1) with replicated observations on non-stochastic x. They studied the relative performance of the OLS predictor and a consistently estimated predictor in terms of an integrated mean square error of prediction (IMSE), where the integration is made over a range of weighted possible values on x_p. Yum and Byun (1990) develop IMSE for a multiple regression model with non-stochastic regressors. They compare OLS with two consistent estimators and suggest IMSE as a measure of overall, average prediction accuracy when the variables are subject to error.

The idea of this paper is to study the performance of a predictor, $\hat{Y}=a+bX$, of $E(Y|x_p)$ for different fixed values of the true x_p. The predictors to be compared are mainly the OLS predictor and a predictor based on consistent estimation of β. This will be done theoretically and by simulations for model (1). We will also study the problem with unequal error variances of X_t, both in the case of non stochastic and stochastic x.

2. THEORY

2.1. **Estimation of the structural parameters.** The size of the asymptotic bias *plim*$(b - \beta)$ of the OLS estimator (b_{OLS}) of β in (1) is related to the so called *reliability ratio*, K_{OLS}, according to:

$$plim\, b_{OLS}= \beta K_{OLS} \tag{2}$$

where $b_{OLS} = \dfrac{S_{XY}}{S_{XX}} = \dfrac{\Sigma(X_t - \overline{X})(Y_t - \overline{Y})}{\Sigma(X_t - \overline{X})^2}$

and where $K_{OLS}=\sigma^2_x/(\sigma^2_x+\sigma^2_u)$.

There are several ways of correcting for such bias (see e.g. Fuller, 1987). If K_{OLS} is known, an evident estimator of β is b_{OLS}/K_{OLS}. For known error variance in X_t Fuller (1987, p. 193) suggests the following consistent moment estimator of β in (1):

$$b_{FUL} = \frac{S_{XY}}{S_{XX} - (1 - \delta / T)\Sigma \sigma^2_{ut}} \qquad \text{if } \lambda^* = \lambda - 1/T \geq 1, \tag{3a}$$

and

$$b_{FUL} = \frac{S_{XY}}{S_{XX} + \delta/_T \, \Sigma \sigma^2_{ut} - (\lambda - 1/_T) \Sigma \sigma^2_{ut}} \qquad \text{if } \lambda^* = \lambda - 1/T < 1, \tag{3b}$$

where $\lambda = S_{XX}(1 - R^2_{XY})/\Sigma\sigma^2_{ut}$.

In (3) δ is a positive constant, that will simply be set to one in this study. λ is the ratio between the residual sum of squares, when taking the reverse regression of X_t on Y_t, and the sum of error variances in X_t. This ratio is expected to be greater than or equal to one since the error variance in X_t is a part of this residual sum of squares. However, in a sequence of observations, it can happen that $\lambda<1$. This is a reason for the modification in (3b), also excluding the possibility of a negative denominator. As the denominator in (3) is bounded below by a positive number, the estimator has moments. For further discussion of this estimator see Fuller (1987) and Jonsson (1992).

If the error variance σ^2_{up} of X_p is different from the mean error variance during the estimation period, then we propose the following modified OLS estimator:

$$b_{MOLS} = \frac{S_{XY}}{S_{XX} - (1 - \delta / T)(\Sigma \sigma^2_{ut} - T\sigma^2_{up})} \,, \qquad \text{if } \lambda^* = \lambda - 1/T \geq 1. \tag{4}$$

If $\lambda^{*}<1$, a modification corresponding to that in (3b) is made. In estimator (4) we correct for the discrepancy in error variance between the estimation and prediction period. This modified OLS estimator (MOLS) gives an estimate of β corresponding to that which would have been obtained by OLS if the error variance of X_t had been the same in the estimation period as in the prediction period. Hence the probability limit of (4) is

$$plimb_{MOLS} = \beta \frac{\sigma_x^2}{\sigma_x^2 + \sigma_{up}^2}. \tag{5}$$

If x_t is random and normally distributed, (4) will yield consistent predictions of Y given observed X_p. The proof follows immediately from that of Johnston (1972, pp. 290, 291) for the case of equal error variances.

The estimate, a, of α is obtained from $a = \overline{Y} - b\overline{X}$.

2.2 MSE for a linear predictor of $E(Y|x_p)$. Let us use model (1) for prediction and let x_p be a fixed number. Then the prediction error for any predictor $\hat{Y} = a + bX_p$ of $Y|x_p$ is

$$Y|x_p - \hat{Y} = \alpha + \beta x_p + q_p + w_p - a - bX_p, \tag{6}$$

where a and b are estimates of α and β obtained from the estimation period sample and hence independent of q_p, w_p and X_p. Taking the squared expectation of (6) yields

$$E(Y|x_p - \hat{Y})^2 = E[E(Y|x_p) - \hat{Y}]^2 + \sigma^2_q + \sigma^2_w. \tag{7}$$

The expectation of the squared prediction error of $y|x_p$ is obtained from (7) by letting $\sigma^2{}_w$=0. The choice of predictor $\hat{Y}$ only affects the first term. Let us therefore in the sequel study MSE of the predictor of E($Y|x_p$)=E($y|x_p$). Hence the prediction error now is:

$$\mathrm{E}(Y/x_p) - \hat{Y}_p = \alpha + \beta x_p - a - bX_p. \tag{8}$$

Denote *plim a* with α^* and *plim b* with β^* and assume that the estimators have moments. Taking the MSE of the predictor over X_p, given x_p yields:

$$\begin{aligned}
MSE_T &= \mathrm{E}(\alpha + \beta x_p - a - bX_p)^2 \\
&= \mathrm{E}(\alpha + \beta x_p - \alpha^* - \beta^* X_p + \alpha^* + \beta^* X_p - a - bX_p)^2 \\
&= (\alpha + \beta x_p - \alpha^* - \beta^* x_p)^2 + \beta^{*2}\sigma^2_{up} && \text{(A+B)} \qquad (9)\\
&+ \mathrm{E}(\alpha^* + \beta^* X_p - a - bX_p)^2 && \text{(C)}\\
&+ 2\,\mathrm{E}(\alpha + \beta x_p - \alpha^* - \beta^* X_p)(\alpha^* + \beta^* X_p - a - bX_p). && \text{(D)}
\end{aligned}$$

The first two terms (A and B) are constants. The third term (C) can be rewritten as:

$$C = \mathrm{E}(\alpha^* + \beta^* x_p - a - bx_p)^2 + \mathrm{E}(\beta^* - b)^2\sigma^2_{up}. \tag{10}$$

C is under general conditions of order T^{-1} (see e.g. Cramér, 1946, pp. 353-354). Rewriting the cross product term D in (9) yields:

$$\mathrm{D} = -2\beta^*\sigma^2_{up}\,\mathrm{E}(\beta^* - b) + 2(\alpha + \beta x_p - \alpha^* - \beta^* x_p)\left[\mathrm{E}(\alpha^* - a) + x_p\,\mathrm{E}(\beta^* - b)\right]. \tag{11}$$

Since the expectations E(α^*-a) and E(β^*-b) are of order T^{-1}, also D is of order T^{-1}. The A and B terms are therefore, at least for large samples, likely to dominate MSE, unless $\sigma^2{}_{up}$ is close to zero. It is worth mentioning this can easily be extended to the case of several x variables. The term B in (9) then becomes $E\{(\Sigma \beta_i^* u_{ip})^2\}$, where the summation extends over all x variables.

Let us focus on the sum of the first two terms in (9), A+B, that can be expressed as:

$$MSE_\infty = \beta^2\{(K-1)^2(\mu_x - x_p)^2 + K^2\sigma_{up}^2\} \quad \text{where } K = \beta^* / \beta . \tag{12}$$

If the predictor is based on OLS, then K is equal to K_{OLS} in (2), and for a consistent estimator of β the ratio K is equal to one. The minimum of (12) occurs for

$$K = (\mu_x - x_p)^2 / [(\mu_x - x_p)^2 + \sigma_{up}^2]. \tag{13}$$

Let $x_p = \mu_x + c\sigma_x$, where μ_x is the mean and σ^2_x is the variance of x in the estimation period. K then becomes

$$K = c^2\sigma_x^2 / [c^2\sigma_x^2 + \sigma_{up}^2]. \tag{14}$$

For equal error variances, $\sigma^2_{up}=\sigma^2_u$ OLS gives the minimum of MSE_∞, if the observed X_p comes from a true x_p at one standard deviation distance from the mean (c=1). K is equal to one only if $c = \infty$, unless σ^2_{up} is zero. Hence a predictor based on consistent estimation of α and β is never optimal, unless x_p is measured without error. When the error variance in X_p is different from that of the estimation period, a predictor based on MOLS according to (4) will be optimal for c=1.

Now compare MSE_∞ of OLS with MSE_∞ of a consistent estimator:

$$MSE_{\infty OLS} - MSE_{\infty CONS} = \beta^2(K_{OLS}-1)\{(K_{OLS}-1)(\mu_x - x_p)^2 + (K_{OLS}+1)\sigma_{up}^2\}. \tag{15}$$

Letting $\sigma_{up} = r\sigma_u$ and replacing σ_{up}^2 in (15) with $\sigma_{up}^2 = r^2\sigma_x^2(1-K_{OLS})/K_{OLS}$ yields

$$MSE_{\infty OLS} - MSE_{\infty CONS} = \beta^2(K_{OLS}-1)^2\{(\mu_x - x_p)^2 - r^2(K_{OLS}+1)\sigma_x^2/K_{OLS}\}. \qquad (16)$$

For a predictor based on a consistent estimator to be better than the OLS predictor in terms of MSE_∞ it must hold

$$(\mu_x - x_p)^2 > r^2\sigma_x^2(K_{OLS}+1)/K_{OLS} \qquad (17a)$$

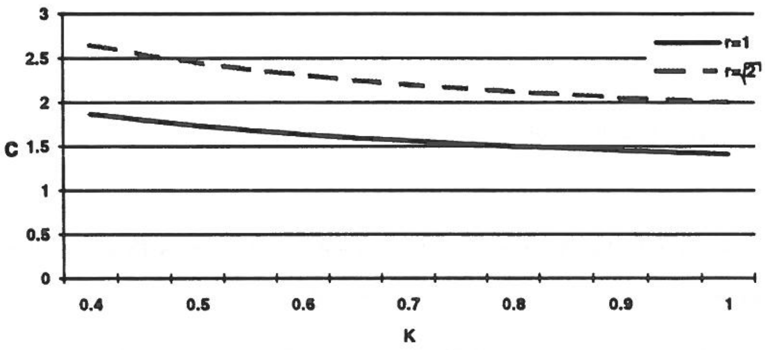
i.e.

$$\left|\mu_x - x_p\right| > r\sigma_x\sqrt{1+\tfrac{1}{K_{OLS}}}\text{ , where } 1+\tfrac{1}{K_{OLS}} \geq 2 \text{ since } K_{OLS} \leq 1. \qquad (17b)$$

The interpretation of the inequality is: If X_p comes from an x_p with a deviation from μ_x smaller than the right hand side of (17b), the OLS predictor should be preferred to a predictor based on a consistent estimator. The expression (17) also tells us that the gain from using OLS is increasing with decreasing K_{OLS}. One faces the paradox that the more important it is to adjust for measurement error, when estimating the structural parameters, the more important it is to use OLS for prediction. This is so although the bias of the OLS predictor increases as K_{OLS} decreases. If the error variance in X_p is greater than the mean error variance of the estimation period, the relative gain from OLS increases. If the error variance in the prediction period is the smaller one, the opposite holds and an extreme case occurs when we observe the true x_p. In that situation the optimal K is equal to one.

In Figure 1, the right hand side of (17b) with σ_x=1 has been plotted against K_{OLS} for r=1 and for $r=\sqrt{2}$. This gives the number of standard deviations x_p has to deviate from μ_x in order for the predictor based on consistent estimation to produce better predictions than OLS. For instance, if K is equal to .8 and r is equal to one the prediction x_p must lie at least one and a half standard deviation from the mean for the predictor based on consistent estimation to be better than the OLS predictor. When the error variance in X_p is twice the error variance in the estimation period, x_p must deviate with at least two standard deviations for this to be the case.

FIGURE 1. Distance c (number of standard deviations) from μ_x that x_p has at least to be in order for the OLS predictor to be worse than a "consistent" predictor. $T=\infty$.



Economic time series are often cyclical or/and contain trends. Let us therefore take a closer look at these two situations under the assumption that x is non-stochastic and start with the cyclical case. If $x_t=\sin(\pi t/s)$ $t=1...2s$, the variance of x_t is .5 and the maximum of x_t minus μ_x is equal to one. The minimum of the right hand side of (17b) is one for r=1, which means that a predictor based on a consistent estimator is never to be preferred to the OLS predictor if r≥1.

Assume in the trend case that x_t, $t=1..T$, is equally spaced between $\min(x_t)$ and $\max(x_t)$[1]. Then the variance of x_t is:

[1]The probability limits of a and b then concern those obtained assuming a large number of independent replications of the measurements on $\{x_1, ..., x_T\}$.

$$\sigma_x^2 = \frac{T+1}{3(T-1)}\left[\max(x_t) - \mu_x\right]^2 . \tag{18}$$

Inserting this into (17a) gives

$$(x_p - \mu_x)^2 > r^2 \frac{T+1}{3(T-1)}\left[\max(x_t) - \mu_x\right]^2\left(1+\tfrac{1}{K_{OLS}}\right). \tag{19}$$

Now assume that x_p=max(x_t). It is then easily seen that the inequality does not hold for r=1 and K_{OLS}≤.5 or for all K_{OLS} if r^2≥1.5. Therefore OLS in these situations is always to be preferred given min(x_t)≤x_p≤max(x_t). Next assume that $x_t=mt$, where m is a positive constant, that the error variance in X_t is proportional to the level of x_t and that we want to make predictions for the next observation, x_{T+1}. Then $r=\sqrt{2}$ and it is possible to show that the OLS predictor is always better than one based on a consistent estimator. Of course, this is the case for all $\sigma^2_{up}=\sigma^2 x^m$ when m≥1. It is also worth mentioning that there are other predictors that sometimes can do better than both the OLS and the consistent estimator based predictor. An example is the cyclical case where the minimum of (14) occurs for the maximum of x_p when c^2=2. In the trend case the minimum occurs for X_{T+1} when c^2≅3.

An alternative predictor in the case of unequal error variances is the MOLS predictor based on estimation of β according to (4). The relationship between this predictor and the one based on consistent estimation is obtained from (17) by letting r=1 and replacing K_{OLS} with K_{MOLS}. This means that all results from a comparison of a predictor based on consistent estimation with the OLS predictor in the case of equal error variances carry over to a comparison with the MOLS predictor in the case of unequal error variances. One interesting special case occurs when x is measured without error in the estimation period, but an estimate of x_p is used when making predictions. Then OLS is a consistent estimator of β. Hence the MOLS predictor will be better than the OLS predictor for a wide range of values on x_p and an OLS prediction is often not optimal in such a case. If the error variance is larger in the prediction period than the mean of the error variances in the estimation period, it can be shown that OLS is to be preferred to MOLS if

$$(x_p - \mu_x)^2 > \frac{K_{OLS} + K_{MOLS}}{2-(K_{OLS} + K_{MOLS})} \frac{1-K_{MOLS}}{K_{MOLS}} \sigma_x^2 \ . \tag{20}$$

Since the MOLS predictor is optimal when x_p deviates by one standard deviation from the mean, the MOLS predictor will always be best if x_p lies within one standard deviation from the mean. If the error variance of X_p is smaller than the mean error variance of the estimation X_t, the MOLS predictor has smaller MSE_∞ than the OLS predictor if (20) holds. The MOLS predictor will therefore always be a better choice if x_p deviates by one standard deviation or more from the mean.

The reason for the relatively better performance of the OLS predictor than one based on consistent estimation is the smaller B term in (9), $K^2\beta^2\sigma^2{}_{up}$. This often compensates the bias of the OLS predictor. If possible, the term B should be reduced. For instance, if the true x_t can be assumed to have a cyclical pattern, then the changes in the values from one observation to the next can be assumed to be small, and it seems reasonable to smooth the observed X values in the prediction period in order to reduce $\sigma^2{}_{up}$. The cost can be a greater bias, but this cost can be smaller than the gain. The MOLS predictor then may be an alternative to the OLS predictor in accordance with the earlier discussion.

Above we have assumed large samples so the C term and D term in (9) could be disregarded. However, we know that at least the C term should be smaller for the OLS predictor than for a predictor based on a consistent estimator because of a smaller variance of the OLS estimator of β. A conjecture is, if we add this to the earlier results, we will get a wider area of values of x_p for which OLS provides the better predictor. The gain for OLS in the C term also increases with decreasing K_{OLS}. If the error variance in X_p is greater than the error variance in the estimation period, the MOLS estimator of β may have a smaller variance than OLS, possibly resulting in a relatively smaller C term.

3. A SIMULATION STUDY

3.1 **Introduction.** In this section some simulation experiments will be performed to illustrate the earlier discussions and to compare different predictors in the case of small samples. The model to be used to generate observations on $\{Y_t, X_t\}$ is model (1) with β=1. In the first set-up x_t is normally distributed. In the other experiments x_t will be non-stochastic. The error variance in X_t is assumed to be known. The explanatory power of the true regressor[2] is set to .8 in all experiments. The predictors to be studied are mainly the OLS predictor, the MOLS predictor and the predictor based on the consistent estimator (3) of β. The last predictor will be called the Fuller predictor. These predictors will be used also in cases of unequal error variances although other predictors based on weighted regression could have smaller variance. The evaluation of the predictors in the cases of non-stochastic x_t will be done by comparing bias and root mean square error (RMSE) at the prediction of $E(Y|x_p)$. Bias and RMSE for the predictors have been calculated from the following relationships assuming the error in the prediction period to be independent of the errors in the estimation period.

The prediction error is

$$E(Y / x_p) - \hat{Y} = \alpha + \beta x_p - a - bx_p - bu_p. \tag{21}$$

---

[2] $R^2_{Yx} = 1 - (\sigma^2_q + \sigma^2_w)/\sigma^2_Y$

Taking the expectation of (21) over u_p yields

$$Bias = \mathrm{E}(\alpha + \beta x_p - a - bx_p). \tag{22}$$

MSE of the predictions can easily be shown to be

$$MSE = \mathrm{E}(\alpha + \beta x_p - a - bx_p)^2 + \mathrm{E}(b^2)\sigma^2_{up}. \tag{23}$$

For each x_p, the bias can be estimated by the mean of the simulated prediction errors, $\alpha + \beta x_p - a - bx_p$. An estimate of MSE is obtained by averaging the squared errors and the squared slope estimates and then using (23) with known error variance.

3.2 **Random x_t.** The first simulation experiment is performed to illustrate the behaviour of the predictors when x_p is a random selection from the same population that generated x_t in the estimation period sample. x_t is assumed to be normally distributed with zero mean and variance equal to .8. In the estimation period sample σ^2_u is equal to .2 and in the prediction period σ^2_{up} is set to .2 in a first set-up and then changed to .4 in a second set-up. The size of the estimation period sample is T=50. The number of replications is 500. The simulation results for the estimators of β are given in Table 1. We note that the means of the estimates are of expected sizes, 1 for Fuller, .8 for OLS and .667 for MOLS, and that RMSE for the Fuller estimator is much smaller than for the other estimators. The number of replications when λ^* in (3) is less than one is only five. Hence there are few such "bad" samples. For a discussion of these matters, see Jonsson (1992).

TABLE 1. Some results for the estimators of β=1.

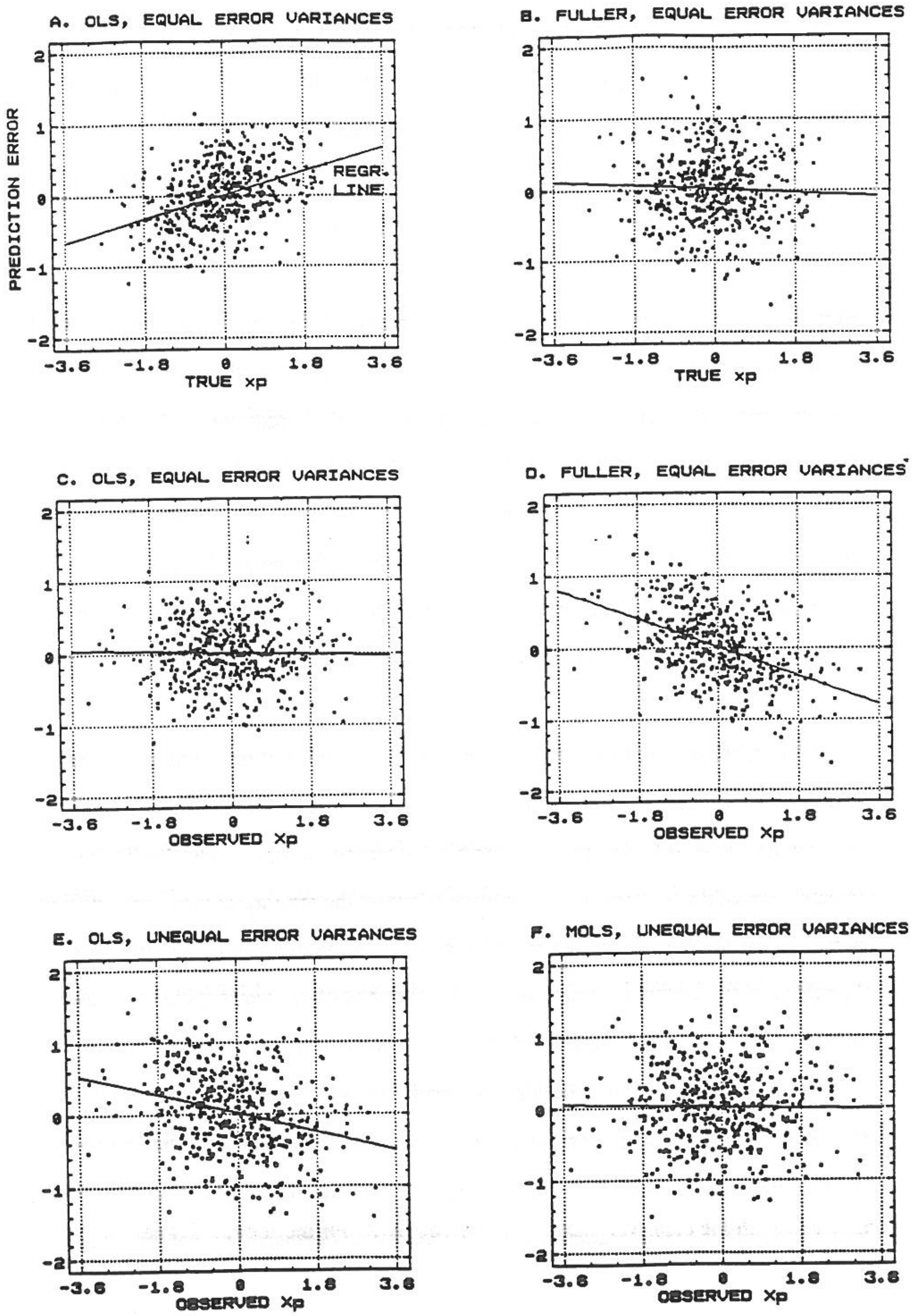
	OLS	FULLER	MOLS, σ^2_{up}=.4	No of repl. $\lambda^*<1$
Mean(b)	.795	1.004	.661	5
RMSE	.223	.133	.347	

The results from the prediction evaluation are summarised in Figure 2. In Figures 2A and 2B the prediction errors according to (21), obtained with the OLS predictor and the Fuller predictor, are plotted against x_p for the situation when σ^2_{up}=.2. The Fuller predictor gives at least approximately unbiased predictions of $E(Y|x_p)$. This is not the case for OLS. In Figures 2C and 2D the prediction errors are plotted against observed X_p. The Fuller predictor yields biased predictions of $E(Y|X_p)$ and now the OLS predictions seem to be unbiased. It can also be mentioned that RMSE for the 500 predictions is somewhat smaller for OLS than for the Fuller predictor.

Figures 2E and 2F present results for the situation when the error variance in the prediction period is assumed to be twice the error variance in the observation period. The compared predictors are the OLS and the MOLS predictors. We note that the OLS predictor produces biased predictions of $E(Y|X_p)$, unless X_p is close to its mean, while the MOLS predictor, that takes account to the change in error variance, gives approximately unbiased predictions for all X_p. RMSE for the MOLS predictor is only marginally smaller than for OLS, when calculated over all observed X_p, but the gain with the MOLS predictor increases with the distance of X_p from its mean.

3.3 **Non-stochastic and cyclical *x***. In this section x_t will be treated as non-stochastic and cyclical according to x_t=sqrt(1.6)sin(πt/12). The estimation period consists of two whole cycles (T=48) and the prediction period of one cycle (t=49:72). This makes it possible to evaluate the predictors when the true x_p comes from different parts of a cycle. When x_t has a cyclical pattern we often have an idea based on earlier X_t-values of the cyclical position of x_p. The variance of x_t over a cycle is .8. σ^2_u will first be set to .2 and in a second set-up to .4, which means that K_{OLS} will be .8 and 2/3, respectively. A trial will also be made by smoothing the observed X_p by successively taking the average of the last two observations of X_p. Bias and RMSE are calculated from (22) and (23) with x_p replaced by the corresponding smoothed value and with the error variance in X_p divided by 2. Furthermore an attempt is

FIGURE 2. Prediction errors of the OLS predictor, the Fuller predictor and the MOLS predictor plotted against true x_p and observed Xp when $\sigma^2_{up}=\sigma^2_u$ (equal error variances) and when $\sigma^2_{up}=2\sigma^2_u$ (unequal error variances).



made to see how the predictors behave if the cycle under the prediction period has a greater amplitude. This will be done by letting the variance of x_t in that cycle be twice the variance during the estimation cycles. The number of replications is set to 500.

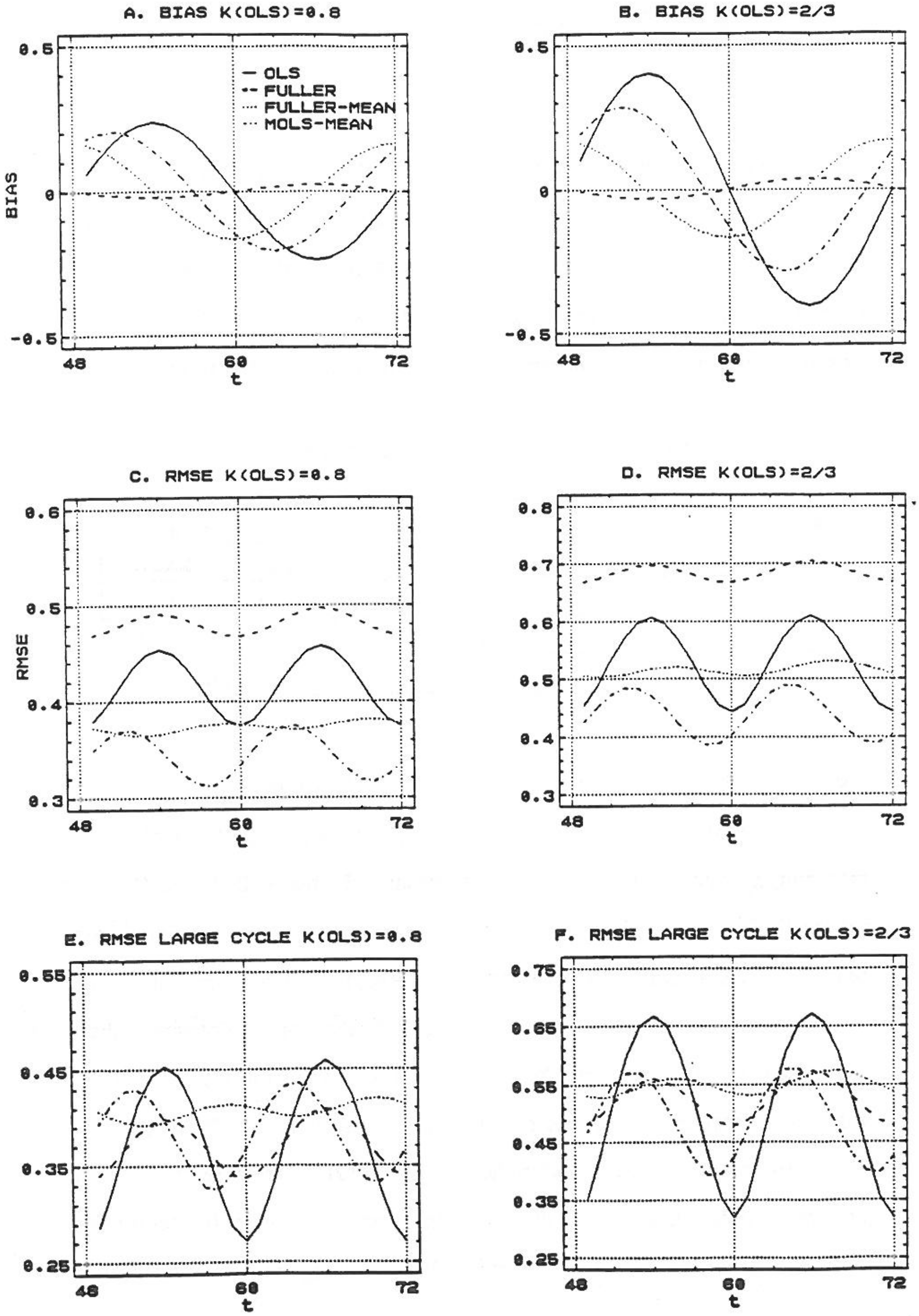
Results from the estimation of β are given in Table 2. The estimates of β are on average somewhat greater than the corresponding probability limits (=K, as β=1). RMSE for the consistent estimator is about 60% of RMSE of the OLS estimator when K_{OLS}=.8. The corresponding number for K_{OLS}=2/3 is 50%. Hence, when estimating the structural parameters the consistent estimator outperforms OLS.

TABLE 2. Some results from the estimations of β=1.

	K_{OLS}=0.8		K_{OLS}=2/3	
	Mean b	RMSE	Mean b	RMSE
OLS	.813	.19947	.681	.3294
FULLER	1.018	.1243	1.027	.1687
No of repl. with λ^*<1	7		66	

Results from the prediction evaluation are presented in Figure 3. Starting with the results for the OLS predictor (OLS) and Fuller predictor (FULLER), applied without smoothing X_p, and in the case of equal amplitudes (figures A, B, C and D), we note that the OLS predictor yields biased predictions while the Fuller predictor produces approximately unbiased predictions. However, RMSE is always smallest, even if we know that we are at the top of a cycle, for the OLS predictor. The picture is the same for K_{OLS}=.8 and K_{OLS}=2/3, but the gain with OLS is larger for K_{OLS}=2/3. All these results are expected according to the discussion in Section 2.2. The reason for the better performance of OLS is its smaller B term in (9). In the case of smoothed X_p this term will be reduced by 50%, while the bias term will only marginally be affected for both the OLS and the Fuller predictor, making the Fuller predictor more

FIGURE 3. Bias and RMSE for different predictors when x_t is cyclical.



competitive. The results presented in the smoothed case are for the Fuller predictor (FULLER-MEAN) and the MOLS predictor (MOLS-MEAN). As can be noted, the prediction errors of the Fuller predictor are much reduced, but the MOLS predictor mostly yields better predictions for K_{OLS}=.8 and always for K_{OLS}=2/3. Smoothing the X_t values also during the observation period in combination with OLS (not in the figure) gives almost identical results as for the MOLS predictor on smoothed X_p. It is also worth mentioning that the OLS predictor, when only X_p was smoothed, produced a somewhat larger RMSE on the average than the MOLS predictor. Only when x_p was close to the mean of x_t, OLS provided a better predictor. These results, too, are consistent with those of Section 2.2.

Figures 3E and 3F show the results for the case when the cycle during the prediction period has a greater amplitude. In the case of non smoothed X_p the Fuller predictor is now to be preferred to the OLS predictor at the top and at the bottom of the cycle, while the OLS predictor is still better elsewhere. It can also be noted that the potential gain from smoothing can be questioned in this case.

3.4 Non-stochastic x_t with a trend. In this section the trend case will be studied by letting x_t be equal to t. The error variance in X_t is first assumed to be constant and determined so that K_{OLS}=.8. In a second experiment the error variance is assumed to be proportional to x_t, $\sigma^2_{ut}=\sigma^2 x_t$. In many situations this is a more realistic assumption than that of a constant error variance. The estimation of the parameters will be performed for two sample sizes, T=50 and T=200. Predictions will then be made when T=50 for x_p=26...70 and when T=200 for x_p=101..230. The within sample predictions makes it possible to study exactly when a predictor is better than another. The number of replications is set to 1000. The estimation results are found in Table 3. The estimators seem to have a small positive bias for T=50. The consistent estimator is superior to the OLS estimator in terms of RMSE. This is especially the case for T=200.

TABLE 3. Some results from the estimations of β=1.

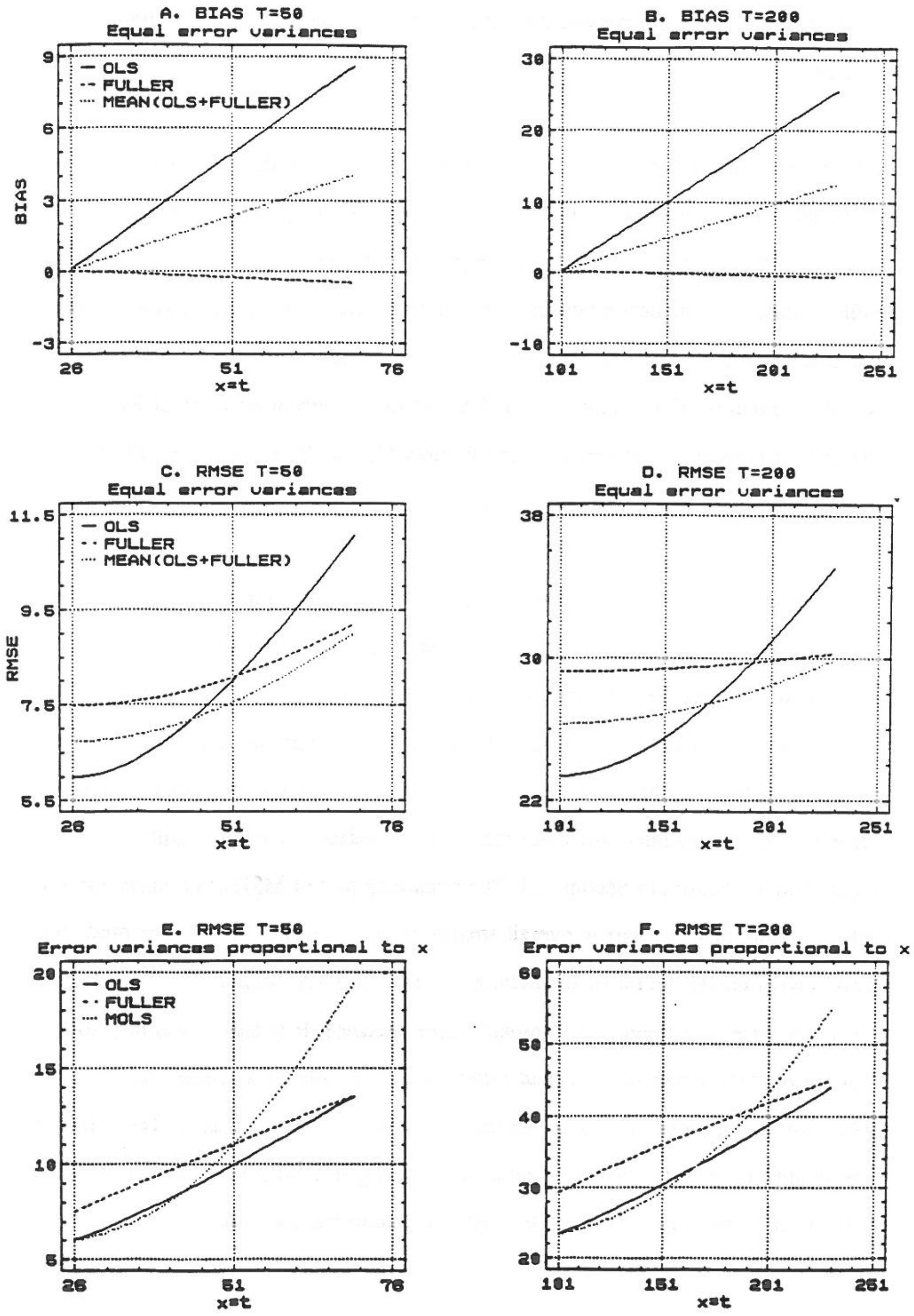
	Equal error variances				Unequal error variances			
	T=50		T=200		T=50		T=200	
	Mean b	RMSE	Mean b	RMSE	Mean b	RMSE	Mean b	RMSE
OLS	.807	.209	.803	.201	.807	.210	.803	.201
FULLER	1.011	.122	1.005	.060	1.011	.124	1.005	.063
No repl. $\lambda^*<1$	10		0		14		0	

In Figures 4A-D bias and RMSE obtained in the prediction evaluation are plotted against $x_p=t$ in the case of constant error variance. This is done for the OLS predictor and the Fuller predictor and the mean of these two. As can be noted, the bias is large for the OLS predictor when x_p lies far away from the mean of x_t in the estimation period, while the bias for the Fuller predictor is always of negligible size.

Looking at Figure 4C, we note that RMSE of the OLS predictor is smaller than RMSE of the Fuller predictor up to a x_p equal to 52. Hence OLS provides the better predictor even for a value of x_p outside the range of the x_t values used for estimation. This is not the case for T=200. The predictors then do equally well when x_p is about 190. Thus the OLS predictor is the one to be preferred of the two predictors for T=50, unless the prediction is made for a long forecast horizon, while the Fuller predictor is to be preferred for T=200, also for a short horizon. As can also be noted from Figures 4C and 4D, there is at least one other predictor that is better than both the OLS and Fuller predictors for values on x_p that lies around T, namely the mean of the two predictors.

RMSE for the case with error variances in x_t proportional to x_t is presented in Figures 4E and 4F. The OLS predictor is now always better than the Fuller predictor over the studied range of x_p values . Furthermore it can be noted that the MOLS predictor is behaving as the OLS predictor in the case of equal error variances. If instead we

FIGURE 4. Bias and RMSE for different predictors when $x_t=t$. Estimation periods are for $T=50$ and $T=200$. Predictions are made from $x_p=26$ and $x_p=101$, respectively.

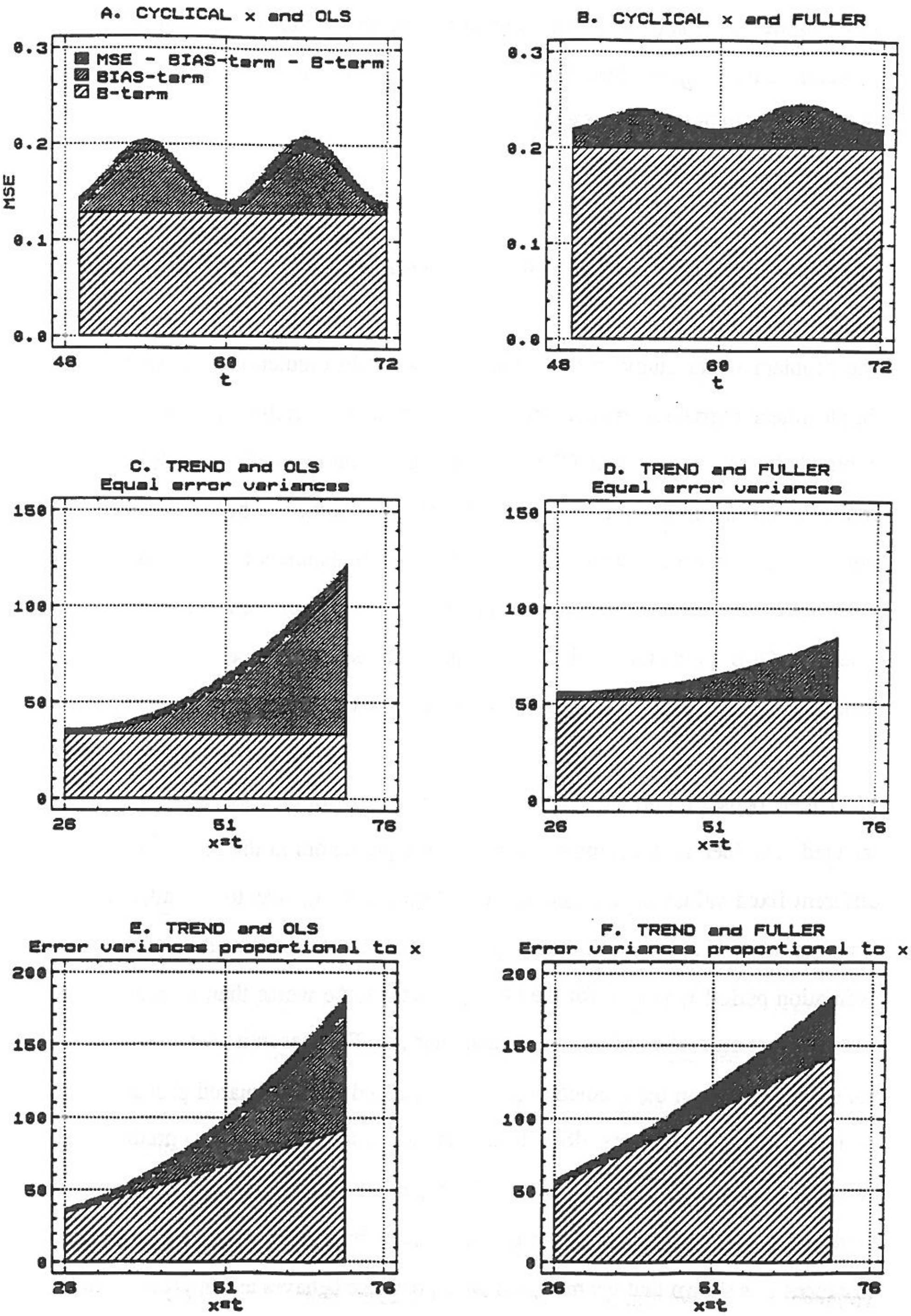


assume a negative trend in x_t (>0), this implies an error variance of X_{T+1} that is less than the mean error variance of X_t in the estimation period. The choice will then be between the MOLS predictor and the Fuller predictor, at least if the estimation sample is large.

3.5 **Decomposition of MSE.** Let us in some of the situations studied above decompose MSE into squared bias (A), the term B ($\beta^{*2}\sigma^2_{up}$) and the remaining part of (9) (C+D). The predictors that are compared in the analysis are the OLS and the Fuller predictors. The decompositions are for the cyclical case with K_{OLS} equal to .8 presented in Figures 5A and 5B. The results for the trend case with T=50 are to be found in Figures 5C-F. Figures 5C and 5D show the decompositions of RMSE for the case of constant error variance, and Figures 5E and 5F show the results for the case with error variance proportional to x_t.

For all situations the results show that MSE of the OLS predictor (Figures 5A, 5C and 5E) is dominated by the bias term plus the B term and MSE of the Fuller predictor by the B term (Figures 5B, 5D and 5F) . We note for the OLS predictor that the bias term in the cyclical case is never large enough to compensate the larger B term of the Fuller predictor. The opposite holds for large x_p in the trend case when a constant error variance is assumed All these results are consistent with the results from the theoretical derivations in Section 2.2. The remaining part of MSE often plays a minor role. We note that this part is overall smaller for OLS than for the Fuller predictor. This is what makes the Fuller predictor better than an OLS predictor only from t=52 on, in the case of a trend and a constant error variance. It is also worth mentioning that the remaining part of MSE had probably been smaller if a greater value on the R^2_{Yx} had been chosen or if larger estimation samples had been used. The effect is presumably largest for the Fuller predictor, making it a little bit more competitive with the OLS predictor. However, this will not change the main results.

FIGURE 5. Decomposition of MSE when x_t is cyclical (K_{OLS}=0.8) and $x_t=t$ (T=50).



A situation when the sum of the C and D term can constitute a large part of MSE occurs when we want to make predictions conditional on a true x_p. The B term is then zero and we note that the Fuller predictor is better or much better than the OLS predictor, unless x_p is close to the mean of x_t. See also Stanley (1988) for a simulation study in the case of known x_p.

4. SUMMARY

The problem under study in this paper is how to do predictions on the basis of a simple linear regression model when the independent variable x_t is measured with errors. It is well-known that OLS, conditional on observed X_t provides consistent predictions if the prediction value X_p on X_t can be assumed to be a random selection from the same normal distribution that generated the estimation period sample. This is not the case if the error variance in X_p differs from that of the estimation period. A modified OLS predictor, which takes this into account, based on known error variances, is proposed and can be shown to produce consistent predictions.

The paper focuses on predictions when the true x_p cannot be assumed to be randomly selected. The idea is to compare MSE of some predictors in the form $\hat{Y}=a+bX_p$ for different fixed values on the true x_p. It is shown that x_p has to lie rather far away from the mean of x_t in the estimation period, sometimes even outside the range of the estimation period x_t values, for the OLS predictor to be worse than a predictor based on consistent estimation of the slope parameter (β). The reason is that a term, equal to the error variance in the prediction period multiplied by the squared probability limit of the estimator of β, enters MSE. That term can be much larger for a predictor based on consistent estimation than for the OLS predictor which will often make the OLS predictor the best one even if it produces a large bias. In the case of unequal error variances it is shown that the modified OLS predictor behaves as the OLS predictor in the case of equal error variances. Some guidelines are also given when to use the Modified OLS predictor instead of the OLS predictor. One special case occurs when x_p is measured with and x_t in the estimation period sample is measured without errors. Then the OLS predictor provides unbiased predictions, but the Modified OLS predictor will yield better predictions in terms of MSE for a wide range of values on x_p.

Special investigations have been made for the cases when x_t is cyclical and when x_t follows a trend. When x_t is cyclical with constant amplitude, it can under certain conditions be shown that the OLS predictor is better than a predictor based on consistent estimation. Smoothing the X values used for prediction can in such a situation be worth trying in order to reduce the prediction error. In a simulation study the modified OLS predictor in most cases produced better results than the consistent predictor applied on smoothed X_p. For the case when x_t follows a linear trend ($x_t = t$) and the error variance is proportional to x_t it was shown that OLS is to be preferred for short forecast horizons. A simulation study showed that this could also be the case for longer horizons when small samples are used. Even when the error variance was constant the OLS predictor did better for a sample size of $T=50$ than the predictor based on consistent estimation when making predictions for $T+1$.

REFERENCES

Aigner D. J. et al (1986), Latent Variable Models in Econometrics, in Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*, Vol. 2, Amsterdam, North Holland.

Cramér, H. (1946), *Mathematical Methods of Statistics*, Princeton University Press.

Fuller, W.A. (1987), *Measurement Error Models*, New York: John Wiley & Sons.

Ganse, R. A., Amemiya, Y., and Fuller, W. A.(1983) Prediction When Both Variables Are Subject to Error, With Application to Earthquake Magnitudes. *Journal of the American Statistical Association* **78**, 761-65.

Hodges, S.D. and Moore, P. G. (1972), Data Uncertainties and Least Squares Regression. *Journal of Applied Statistics*. **21**, 185-95.

Johnston, J. (1972), *Econometric Methods*, 2nd ed. McGraw-Hill, New York.

Jonsson, B. (1992), Sample Based Proportions as Values on an Independent Variable in a Regression Model. National Bureau of Economic Research (sweden) Working Paper 16, 1992.

Madansky, A. (1959), The Fitting of Straight Lines When Both Variables are Subject to Error. *Journal of the American Statistical Association* **54**, 173-205.

Malinvaud, E. (1970), *Statistical Methods of Econometrics*. North-Holland, Amsterdam.

Pierce, D. A. (1981), Sources of Error in Economic Time Series. *Journal of Econometrics* **17**, 305-21.

Reiser, M., Fecso, R. and Chua, M-K. (1992), Some Aspects of Measurement Error in the United States Objective Yield Survey. *Journal of Official Statistics* **8**, pp. 351-75.

Stanley, T. D. (1988), Forecasting from Fallible Data: Correcting Prediction Bias With Stein-rule Least Squares. *Journal of Forecasting* **7**, 103-13.

Wold, H. (1963), On the Consistency of Least Squares Regression. *Sankhya Ser. A*, **25**, 211-15.

Yum, B. J. and Neuhardt, J. B. (1984), Analysis of the Prediction Problem in a Simple Functional Relationship Model. *IIE Transactions* **16**, 177-84.

Yum, B. J. and Byun, J. H. (1990), Analysis of the Prediction Problem with Errors in the Variables. *IIE Transactions* **22**, 73-83.

Sammanfattning

Tidsseriedata är ofta behäftade med fel av olika slag. Mätningarna kan exempelvis vara preliminära eller vara surveybaserade. Skattningar av ekonometriska modeller baseras ofta på tidsseriedata. Ett exempel är när data från surveyer rörande hushållens attityder till den ekonomiska utvecklingen och deras planer om bilinköp används som förklarande variabler i konsumtions- eller investeringsfunktioner. Eftersom dessa variabler är baserade på urval och således mätta med fel, kommer deras parametrar att bli underskattade i absoluta tal om vanlig minsta-kvadrat metod (OLS) används. I Jonsson (1992) studeras mot bakgrund av de svenska hushållens attityder och planer effekten av urvalsfel på skattningen av riktningskoefficienten i en enkel linjär regressionsmodell, när regressorn består av proportioner. Jonsson fann att effekten kunde vara betydande i vissa fall. Estimatorer som korrigerade för den nämnda underskattningen studerades också och befanns ofta fungera väl relativt OLS. Huvudsyftet med att använda attityd och plandata är dock ofta inte estimation utan prediktion och vad gäller då? Om det X-värde för vilket önskas en prognos kan anses vara slumpmässigt draget ur samma population som gav estimationsstickprovet är det känt att prediktion baserad på OLS bör användas. I exemplet ovan är det tveksamt om detta kan antas - tidsserierna har tydliga cykliska mönster och dessutom varierar mätfelsvariansen över tiden. I denna rapport utreds hur olika prediktorer baserade på observerade tidsserier fungerar för olika sanna värden (x_p) på prediktionsvariabeln. Resultaten kan sammanfattas i följande punkter och är i termer av medelkvadratfel:

Konstant mätfelsvarians. En prediktor baserad på OLS-estimation bör föredras framför en baserad på konsistent estimation om inte x_p ligger långt ifrån medelvärdet av x i estimationsstickprovet. Som exempel kan nämnas att om x är cyklisk är OLS alltid att föredra. En simuleringsstudie visar att OLS kan vara att föredra även i fallet med en trend i x, om prediktion görs för nästa observation och antalet observationer ej är stort.

Olika mätfelsvarians i prediktions- och estimationsperiod. Om mätfelsvariansen är störst under prediktionsperioden växer det område för vilket OLS bör föredras framför en konsistent skattad prediktor. En prediktor baserad på estimation där hänsyn till skillnad i mätfelsvarians tas kan då vara ett alternativ.

En paradox. Ju viktigare det är att justera för mätfelet då syftet är estimation, desto viktigare är det att bevara OLS för prediktion. Detta gäller trots ett växande systematisk prediktionsfel för OLS.

Prevoius titles in this serie:

No. 1 Current Account and Business Cycles: Stylized Facts for Sweden by Anders Warne and Anders Vredin. December 1989.

No. 2 Change in Technical Structure of the Swedish Economy by Göran Östblom. December 1989.

No. 3 Mamtax. A Dynamic CGE Model for Tax Reform Simulations by Paul Söderlind. December 1989.

No. 4 The Supply Side of the Econometric Model of the NIER by Alfred Kanis and Aleksander Markowski. November 1990.

No. 5 The Financial Sector in the SNEPQ Model by Lennart Berg. February 1991.

No. 6 Consumer Attitudes, Buying Intentions and Consumption Expenditures. An Analysis of the Swedish Household Survey Data by Anders Ågren & Bo Jonsson. April 1991.

No. 7 A Quarterly Consumption Funtion for Sweden 1979-1989 by Lennart Berg & Reinhold Bergström. October 1991.

No. 8 Good Business Cycle Forecasts - A Must for Stabilization Policies by Lars-Erik Öller. February 1992.

No. 9 Forecasting Car Expenditures Using Household Survey Data by Bo Jonsson and Anders Ågren. February 1992.

No. 10 Forecasting the Business Cycle Not Using Minimum Autocorrelation Factors by Karl-Gustaf Löfgren, Bo Ranneby & Sara Sjöstedt. February 1992.

No. 11 Current Quarter Forecasts of Swedish GNP Using Monthly Variables by Stefan Gerlach. February 1992.

No. 12 The Relationship Between Manufacturing Production and Different Business Survey Series in Sweden by Reinhold Bergström. February 1992.

No. 13 Forecasting the Swedish Unemployment Rate: V A R vs Transfer Function Modelling by Per-Olov Edlund and Sune Karlsson. March 1992.

No. 14 Business Survey Data in Forecasting the Output of Swedish and Finnish Metal and Engineering Industries: A Kalman Filter Approach by Markku Rahiala and Timo Teräsvirta. March 1992.

No. 15 The Relationship Between Manufacturing and Various BTS Series in Sweden Illuminated By Frequency and Complex Demodulate Methods by Anders Christoffersson, Roland Roberts and Ulla Eriksson. April 1992.

No. 16 Sample Based Proportions As Values On An Independent Variable In A Regression Model by Bo Jonsson. October 1992.

No. 17 Eliciting Turning Point Warnings From Business Surveys by Lars-Erik Öller. November 1992.

No. 18 Volatility, Trading Mechanisms and International Cross-Listing by Margaret M. Forster. December 1992.