# TESTS OF AN ADAPTIVE REGRESSION MODEL

Thomas F. Cooley and Edward C. Prescott *

## Introduction

ANY econometric equation representing a complex behavioral or technical relationship is, of necessity, an approximation of reality. As such, it is subject to errors in specification and structural change over time. This problem is well recognized by econometricians. Duesenberry and Klein (1965) point out that ". . . as technology, institutional arrangements, tastes and managerial techniques change over time, the relationships represented by our equations inevitably change." Furthermore, when statistical tests are applied to econometric relationships, the hypothesis of structural stability is frequently rejected.[1] Some methods for dealing with structural change have evolved. Quandt (1957) has developed a maximum likelihood technique for estimating a point of structural change within a sample.[2] Klein and Evans (1967) adjust the intercepts of the Wharton Model to account for structural change.[3] The purpose of this paper is to test the robustness of Adaptive Regression (1973) to specification errors causing structural change over time, relative to ordinary least squares analysis with and without the autoregressive correction.[4]

Since econometricians are inevitably faced with structural change and errors in specification, they should use a technique which is robust relative to such problems. The device most commonly used is to assume that the disturbances are subject to an autoregressive process. The autoregressive correction may frequently ameliorate the effects of misspecification and structural change, but it is doubtful whether such processes, except in rare instances, describe the true distribution of the disturbances. The economics literature seldom gives any justification for this scheme except that omitted variables may be subject to an autoregressive process or the structure of the model may be changing.[5] We suspect the reasons for the widespread use of the autoregressive correction are that it is a simple hypothesis, explains serial correlation in the disturbances, and can be dealt with efficiently. The adaptive regression model considered in this paper is equally simple but more general, explains serial correlation, and can also be dealt with efficiently.[6]

In the next section the adaptive regression model is presented and the Bayesian estimators are developed. In section II the results of a Monte Carlo Study are presented. Two models are considered for which data are generated by eleven different schemes. The estimation and forecasting efficiency of adaptive regression, and ordinary least squares with and without the autoregressive correction are compared. Section III contains an analysis of the role of time trends in econometric relationships. In section IV the relative forecasting ability of the three estimation techniques is tested on real data. The three models suggested by

Received for publication February 10, 1972. Revision accepted for publication November 30, 1972.

* The authors acknowledge helpful comments of Professors F. G. Adams, R. Roll and R. Summers and the participants of the NBER conference on Bayesian Statistical Inference in Economics. Computations were executed on the University of Pennsylvania computer.

[1] Examples of such tests include Brown (1966), Goldfield (1969) and Howrey (1970). One of the most extensive studies was done by Duffy (1969).

[2] The Quandt technique is limited by the fact that it is mainly useful for finding stable subsamples. If structural change occurs often, it is not very useful. Rosenberg (1968) has used stepwise composition to develop the computationally efficient Aitken estimates of a model subject to structural change over time. His procedure, however, requires that the true covariance matrix of the disturbances be known up to a constant scale factor.

[3] Adjusting the intercepts is an ad hoc method for keeping the model on track for ex ante forecasting. The intercepts are not assumed to change over the sample period which is always much longer than the forecasting period.

[4] The autoregressive correction assumes the error is subject to a first or second order autoregressive scheme. See Dhrymes (1969) for the maximum likelihood approach and Zellner and Tiao (1965) for the Bayesian development. The latter approach is used in this paper.

[5] In fact, if omitted variables are subject to an autoregressive process, the disturbances will, in general, be subject to a more complicated process.

[6] A test with sufficient power to differentiate between these two models (or others which result in serial correlation) using sample sizes generally available to econometricians does not appear to exist. Further, if one did, its usefulness would be limited as neither structure is likely to be an exact representation of reality. That one structure is more likely on the basis of the data does not imply that it will forecast better if, in fact, a third structure is generating the data.

Friedman and Meiselman (1963) are considered. Section V summarizes the findings.

## I An Adaptive Regression Model

The adaptive regression model considered here explicitly assumes that the equation may be subject to permanent structural change over time. The assumed structure is

$$y_t = \beta_{0t} + \beta_1 x_{1t} + \dots + \beta_k x_{kt} + u_t \qquad t = 1,2,\dots,T, \tag{1}$$

where $y_t$ is the dependent variable for period $t$, $\beta_{0t}$ is the intercept value in period $t$, $\beta_i$ the unknown slope coefficients, $x_{it}$ the $t^{\text{th}}$ observation of the $i^{\text{th}}$ explanatory variable, and $u_t$ the additive transitory disturbance.

The intercept is subject to random sequential changes over time:

$$\beta_{0,t+1} = \beta_{0t} + v_t \qquad t = 1,2,\dots,T. \tag{2}$$

The $u_t$ and $v_t$ are assumed to be independent normal variates with mean 0 and variances

$$\text{var}(u_t) = (1-\gamma)\sigma^2 \text{ and } \text{var}(v_t) = \gamma\sigma^2 \tag{3}$$

where $0 \leq \gamma \leq 1$. The unknown parameter $\gamma$ measures the relative importance of permanent and transitory changes. If $\gamma = 0$, all disturbances are transitory and the model reduces to the conventional linear regression model.

Since the process generating the intercepts is nonstationary, it is possible by specifying the value of the intercept at any arbitrary point, to prescribe the joint distribution of any finite set of the other values. As economists are typically interested in forecasting, it is assumed that the objective is to draw inference about the current value of the random intercept. Thus, we set $\beta_0 = \beta_{0,T+1}$ and from (2)

$$\beta_{0,t} = \beta_0 - \sum_{i=t}^{T} v_i, \tag{4}$$

so equation (1) becomes

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i x_{it} + u_t - \sum_{i=t}^{T} v_i. \tag{5}$$

The current value of the random intercept $\beta_0$ is now treated as a parameter. It is clear that this specification bears some similarity to the conventional autoregressive model in which the disturbances are also correlated through time

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i X_{it} + \sum_{i=0}^{\infty} \rho^i e_{t-i}$$

where $e_t$ is assumed to be a normal variate with mean zero and variance $\sigma^2$. This latter structure, however, is less general than the adaptive regression structure because it implies that the effect of omitted factors or disturbances to the process all decay exponentially with time and at the same rate. It is quite conceivable that there may be changes in omitted factors (such as tastes or technology) which will persist into the future without decay. The adaptive regression specification can better capture these effects.

Let $y$ be the $T$ component column vector of the $y_t$, $\beta$ the $k + 1$ component vector

$$\beta' = [\beta_0, \beta_1, \dots, \beta_k] \tag{6}$$

and $X$ the $T$ by $k + 1$ matrix of explanatory variables with $x_{1t} = 1$, $t = 1, \dots, T$. The $TxT$ matrices $R$ and $Q_\gamma$ are defined by

$$r_{ij} = \min[T - i + 1, T - j + 1] \tag{7}$$

and

$$Q_\gamma = (1-\gamma)I + \gamma R. \tag{8}$$

With this notation it is easily verified that

$$y \sim N[X\beta, \sigma^2 Q_\gamma]. \tag{9}$$

The maximum likelihood estimators of the adaptive regression model and their large sample properties were developed in (Cooley and Prescott, 1973). In this paper we present the results of Bayesian estimation of the model. We assume the prior knowledge about the parameters $\beta$, $\sigma$ and $\gamma$ can be represented by locally uniform and independent distributions:[7]

$$\begin{aligned} p(\gamma) &= d\gamma \qquad 0 \leq \gamma \leq 1 \\ p(\beta) &\propto k_1 \\ p(\sigma) &\propto 1/\sigma d\sigma \end{aligned} \tag{10}$$

Let $B_\gamma$ be the Aitken estimator of $\beta$

$$B_\gamma = (X'Q_\gamma^{-1}X)^{-1}X'Q_\gamma^{-1}y, \tag{11}$$

and $S_\gamma$ be the generalized sum of squared residuals

$$S_\gamma = (y - XB_\gamma)'Q_\gamma^{-1}(y - XB_\gamma). \tag{12}$$

Following the analysis of Zellner and Tiao (1965), the parameters $\beta$ have the posterior density

$$p(\beta;y,\gamma,\sigma) \sim N[B_\gamma, \sigma^2(X'Q_\gamma^{-1}X)^{-1}], \tag{13}$$

and the marginal posterior for $\gamma$ is

[7] We are aware of the admissibility problem when diffuse priors are assumed. Nevertheless, this is the assumption commonly used by econometricians. Alternatively, we could have assumed priors which are proper but sufficiently diffuse that the sample information dominated the prior.

$$p(\gamma;y) \propto |Q_\gamma|^{-1/2}|(X'Q_\gamma^{-1}X)^{-1}|^{1/2} S^{-(T-k)/2}. \quad (14)$$

Conditional on $\gamma$ then, the first moment of $\beta$ is simply the Aitken estimator $B_\gamma$.[8] With expression (14) numerical integration is used to obtain the first moment of the posterior for $\beta$

$$E(\beta;y) = \int B_\gamma p(\gamma;y)d\gamma. \quad (15)$$

Other moments of the posterior distribution can be obtained by numerical integration using the result of Zellner and Tiao (1957, p. 773) that

$$p(\beta|\gamma,y) = \text{const}\left\{1+\frac{[\beta-B_\gamma]'X'Q_\gamma^{-1}X[\beta-B_\gamma]}{S_\gamma}\right\}^{-T/2}. \quad (16)$$

This is a form of the multivariate student $t$ distribution.

Because the random parameter $\beta_0$ is generated by a nonstationary process, it cannot be estimated consistently. It has been shown, however, that the estimates of $\gamma$ and $\sigma^2$ are consistent and that the estimates of the random intercept and the slopes are asymptotically efficient. The concern of the present paper is to examine the performance of these estimators in small and moderate sized samples under realistic conditions. As Malinvaud (1966, p. 71) has pointed out, in the evaluation of estimators it is essential to study the sensitivity of their properties to changes in the assumptions of the model to which they are applied.

## II Monte Carlo Tests

The objective of the Monte Carlo analysis was to test the overall accuracy and robustness of the adaptive regression technique (ADR). Since the adaptive regression model considered here assumes a more general form of serial correlation in the residuals it is of primary interest to examine whether applying the adaptive regression technique yields significant improvements over the more commonly used autoregressive correction. It is also of interest to see how adaptive regression performs in the face of specification errors. Three estimation techniques were used on each sample of test data. The accuracy and predictive efficiency of the ADR estimates were compared to estimates generated by ordinary least square (OLS) and to generalized least squares estimates generated under the assumption that the error terms were subject to a first order autoregressive process (AUTO); that is

$$u_t \sim N(\rho u_{t-1},\sigma^2).$$

In order to provide a broad test for robustness, several different structures were used to generate the test data.

The first set of schemes used the adaptive regression model to generate the data. Different values for $\gamma$ were utilized; this varied the relative importance of permanent and transitory changes. In some situations this structure may be a close approximation of reality, but they are possibly few in number. Specification errors will not always result in intercept changes which are identically and independently distributed normal variates with mean 0 as the ADR model assumes. For this reason two other schemes with very different probability laws governing the intercept change were included to determine whether the results are sensitive to the assumptions concerning the intercept changes. One scheme had a small probability of a large change in the intercept in each period while the other had a constant change in the intercept every period. The latter situation would arise if a variable with a time trend had been omitted from the analysis. Specification errors, which necessarily arise because of the need to approximate, may cause shifts in the slope coefficients as well as the intercept. To determine whether this affects the results, a scheme with randomly changing slope parameters was included.

The final set of data generation schemes had no parameter change. In one, all the assumptions of conventional regression theory were satisfied. This model is a special case of the adaptive structure, obtained when $\gamma = 0$. This permits us to analyze what is lost by making the more general assumption that there may be permanent as well as transitory disturbances. The remaining schemes had disturbances which

[8] It should be noted that repeated inversion of $Q_\gamma$ is not necessary. In Cooley and Prescott (1973) a transformation $P$ is developed which does not depend upon the unknown parameters such that $Py$ is normal with mean $PX\beta$ and diagonal covariance matrix. This was important for without this transformation computation costs of adaptive regression would be excessive given current computer technology.

TABLE 1. — STRUCTURES UTILIZED IN THE ANALYSIS *

| | Intercepts | Slopes | Disturbances |
|---|---|---|---|
| *Intercept Change* | | | |
| Adaptive $\gamma = .25$, $\gamma = .50$, $\gamma = .75$, $\gamma = 1.00$ | $\beta_{0,t+1} \sim N[\beta_{0,t}, 16(1-\gamma)]$; $\beta_{0,0} = 10$ | $\beta_1 = 0.4,\ \beta_2 = 0.6$ | $u_t \sim N(0, 16\gamma)$ |
| Constant change | $\beta_{0,t+1} = \beta_{0,t} + 0.4$; $\beta_{0,0} = 10$ | $\beta_1 = 0.4,\ \beta_2 = 0.6$ | $u_t \sim N(0, 16)$ |
| Few large changes | $Pr[\beta_{0,t+1} = \beta_{0,t}] = 0.9$; $Pr[\beta_{0,t+1} = \beta_{0,t} + 20] = 0.05$; $Pr[\beta_{0,t+1} = \beta_{0,t} - 20] = 0.05$ | $\beta_1 = 0.4,\ \beta_2 = 0.6$ | $u_t \sim N(0, 16)$ |
| *No Intercept Change* | | | |
| Independent disturbances | $\beta_0 = 10$ | $\beta_1 = 0.4,\ \beta_2 = 0.6$ | $u_t \sim N(0, 16)$ |
| First order autoregressive disturbances $\rho = 0.3$, $\rho = 0.7$ | $\beta_0 = 10$ | $\beta_1 = 0.4,\ \beta_2 = 0.6$ | $u_t \sim N(\rho u_{t-1}, 16)$ |
| Second order autoregressive disturbances | $\beta_0 = 10$ | $\beta_1 = 0.4,\ \beta_2 = 0.6$ | $u_t \sim N(1.2u_{t-1} - .5u_{t-2}, 16)$ |
| *Slope Change* | $\beta_{0,t+1} \sim N(\beta_{0,t}, 8)$; $\beta_{0,0} = 10$ | $\beta_{1,t+1} \sim N(\beta_{1,t}, .004)$; $\beta_{2,t+1} \sim N(\beta_{2,t}, .004)$; $\beta_{1,0} = 0.4,\ \beta_{2,0} = 0.6$ | $u_t \sim N(0, 8)$ |

* The $u_t$ are independent of each other and of any randomly changing parameter. Changes in parameters are also temporarily independent and independent of each other.

were generated by first and second order autoregressive processes.[9]

The equations which generated the data had two explanatory variables

$$y_t = \beta_{0t} + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t.$$

The $x_{1t}$ were subject to both permanent and transitory changes. These changes were independent normal variates with mean 0 and variances 16 and 4 for the permanent and transitory component respectively. The initial value of $x_{1,0}$ was 30. For equation A in table 2 the $x_{2t}$ were identically and independently distributed normal variates with mean 0 and variance 25. Since many economic series are highly correlated with time, we thought it of interest to determine how results varied with such a variable present. This was accomplished by changing the distribution of $x_{2t}$ to obtain equation B. Instead of being identically and independently distributed, $x_{2t}$ was normal with mean $t$ and variance 5. For both equations A and B, 100 samples of 35 observations were generated by each of the eleven schemes presented in table 1. Each of the three estimation techniques were used to estimate the equations and to forecast the value of the dependent variable in the period subsequent to the sample data.[10] Both Bayesian and maximum likelihood estimators were used. The results were nearly the same so only the Bayesian results are reported.

Although samples of thirty-five observations are in the range frequently encountered by econometricians, many times only a small number of observations are available to estimate the relationship. To determine if results are sensitive to sample size, the experiments with equation A were repeated using samples of only 20 observations. The results of this extensive

[9] The parameters of the second order process were suggested to us by M. McCarthy and T. Sargent as typical of those obtained in economics when second order autoregressive processes are estimates. The adaptive and autoregressive structures are equivalent when either or both $\rho$ and $\gamma$ equal 0, the case of independent disturbances, or when both equal 1.

[10] The optimal forecast when loss is proportional to the mean squared forecast error is the mean of the predictive probability density function (P.P.D.F.). Letting prime symbols denote means of the posterior, the forecast formulae were

$$y'_{T+1} = \beta'_0 + \beta'_1 x_{1,T+1} + \beta'_2 x_{2,T+1}$$

for OLS,

$$y'_{T+1} = \beta'_{0,T+1} + \beta'_1 x_{1,T+1} + \beta'_2 x_{2,T+1}$$

for ADR, and

$$y'_{T+1} = \beta'_0 + \beta'_1 x_{1,T+1} + \beta'_2 x_{2,T+1} + \rho'(y_T - \beta'_0 - \beta'_1 x_1 - \beta'_2 x_2)$$

for AUTO. These are the means of the P.P.D.F. for $y_{T+1}$ for ADR and OLS and approximately the mean for AUTO.

TABLE 2. — SUMMARY [a] OF FORECASTING RESULTS: MEAN SQUARE ERRORS

| Error Generation Scheme | Equation A (35 obs.) | | | Equation A (20 obs.) | | | Equation B (35 obs.) | | |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | AUTO | ADR | OLS | AUTO | ADR | OLS | AUTO | ADR |
| *Intercept Change* | | | | | | | | | |
| Adaptive $\gamma = .25$ [b] | 38.5 | 19.1 | 10.5 | 33.9 | 17.1 | 14.1 | 22.9 | 16.1 | 10.9 |
| $\gamma = .50$ [c] | 69.3 | 26.9 | 12.2 | 31.8 | 22.2 | 17.4 | 39.3 | 27.4 | 18.1 |
| $\gamma = .75$ [c] | 135.8 | 47.0 | 16.4 | 83.7 | 36.2 | 22.8 | 56.8 | 23.1 | 17.1 |
| $\gamma = 1.00$ [b] | 115.1 | 39.5 | 20.8 | 106.2 | 40.5 | 20.3 | 79.6 | 31.3 | 21.0 |
| Constant change [c] | 29.1 | 16.1 | 5.8 | 80.3 | 48.5 | 38.8 | 4.3 | 4.7 | 4.7 |
| Few large changes [c] | 213.8 | 116.6 | 27.5 | 358.4 | 165.3 | 66.6 | 181.6 | 71.5 | 30.7 |
| *No Intercept Change* | | | | | | | | | |
| Independent disturbances | 2.6 | 3.1 | 4.4 | 2.6 | 3.2 | 5.1 | 2.9 | 3.4 | 4.5 |
| First order autoregressive disturbances | | | | | | | | | |
| $\rho = 0.3$ [b] | 10.1 | 4.4 | 4.0 | 105.8 | 36.7 | 28.2 | 7.7 | 3.6 | 3.6 |
| $\rho = 0.7$ [b] | 19.4 | 8.7 | 5.8 | 122.4 | 12.7 | 14.3 | 16.4 | 5.7 | 4.2 |
| Second order autoregressive disturbances [b] | 63.3 | 29.3 | 31.3 | 120.1 | 55.8 | 54.1 | 53.9 | 28.4 | 33.4 |
| *Slope Change* [c] | 149.7 | 55.7 | 23.4 | 76.5 | 47.0 | 32.4 | 71.6 | 39.0 | 24.7 |

[a] The samples consisted of 100 trials.
[b] Indicates fraction of times ADR was better than OLS is significant at the 5 per cent level.
[c] Indicates fraction of times ADR was better than OLS and AUTO is significant at the 5 per cent level.

TABLE 3. — SUMMARY OF ESTIMATION RESULTS: MEAN SQUARE ERRORS

| Error Generation Scheme | Equation A (35 observations) | | | | | | Equation B (35 observations) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coefficient $\beta_1$ | | | Coefficient $\beta_2$ | | | Coefficient $\beta_1$ | | | Coefficient $\beta_2$ | | |
| | OLS | AUTO | ADR | OLS | AUTO | ADR | OLS | AUTO | ADR | OLS | AUTO | ADR |
| *Intercept Change* | | | | | | | | | | | | |
| Adaptive $\gamma = .25$ | .120 | .064 | .034 | .025 | .020 | .017 | .095 | .060 | .047 | .110 | .074 | .047 |
| $\gamma = .50$ | .270 | .088 | .036 | .086 | .030 | .022 | .058 | .031 | .019 | .290 | .180 | .066 |
| $\gamma = .75$ | .370 | .120 | .030 | .046 | .017 | .014 | .078 | .042 | .028 | .300 | .170 | .066 |
| $\gamma = 1.00$ | .340 | .140 | .040 | .130 | .029 | .019 | .190 | .062 | .033 | .470 | .230 | .077 |
| Constant change | .010 | .051 | .019 | .021 | .019 | .018 | .130 | .150 | .150 | .120 | .110 | .067 |
| Few large changes | .700 | .330 | .100 | .320 | .080 | .063 | .330 | .200 | .100 | 1.300 | .610 | .240 |
| *No Intercept Change* | | | | | | | | | | | | |
| Independent disturbances | .011 | .013 | .014 | .017 | .018 | .017 | .009 | .011 | .014 | .007 | .010 | .024 |
| First order autoregressive disturbances | | | | | | | | | | | | |
| $\rho = 0.3$ | .018 | .023 | .029 | .016 | .015 | .014 | .022 | .025 | .030 | .011 | .012 | .029 |
| $\rho = 0.7$ | .071 | .053 | .046 | .037 | .023 | .020 | .066 | .050 | .049 | .050 | .037 | .042 |
| Second order autoregressive disturbances | .064 | .043 | .050 | .079 | .030 | .032 | .160 | .064 | .059 | .081 | .041 | .054 |
| *Slope Change* | .290 | .170 | .083 | .130 | .095 | .089 | .230 | .130 | .090 | .350 | .250 | .120 |

battery of tests are summarized in tables 2 and 3. To conserve space only the mean squared errors are presented for both the forecasts and the estimates.[11] The mean squared forecasting error (MSFE) is computed as the squared difference between the forecast, $y'_{t+1}$, and the optimal forecast with the true values of the parameters known.[12] This reduces the variability of the results and increases the power of the tests. Nonparametric sign tests were performed to test the significance of the differences in

[11] Using the conventional mean squared error loss function, the Bayesian estimates and forecasts are just the means of the posterior and predictive probability density functions.

[12] With the intercept change schemes, the optimal forecasts, given the true parameters, is $\beta_{0,T+1} + \beta_1 x_{1,T+1} + \beta_2 x_{2,T+1}$. When the slope parameters were subject to change, the values of $\beta_1$ and $\beta_2$ in period $T+1$ were used. When there is no intercept change, and the disturbances are independent, the optimal forecast is $\beta_0 + \beta_1 x_{1,T+1} + \beta_2 x_{2,T+1}$. To this $\rho u_{t-1}$ must be added for the first order autoregressive

forecast errors.[13] Entries in table 2 marked with a [b] indicates that the differences in forecasting efficiency between OLS and ADR were statistically significant; [c] indicates that the differences between AUTO and ADR were also significant.

*Significant Findings*

1) When intercept change is present, the mean squared forecasting error (MSFE) is dramatically lower for ADR. The MSFE was as low as one-fifth that of OLS and one-half that of AUTO. Similar improvements are made in estimation efficiency. The mean squared estimation errors (MSEE) are generally significantly lower for ADR.

2) It appears that the ADR results are not sensitive to errors in specifying the probability laws governing this intercept change. With the exception of the constant change in equation B, ADR yields substantially lower MSFE and MSEE than OLS or AUTO when this form of misspecification exists. The reason for the exception is noted in 7) below.

3) When there is no intercept change, and disturbances are independent, the loss involved in using ADR is quite small. This is a reflection of the fact that the ADR technique is quite accurate at estimating the true value of $\gamma$.

4) Somewhat surprisingly, ADR performed as well or better than AUTO in terms of mean squared forecasting error when the disturbances were generated by a first order autoregressive process. These differences were not statistically significant, however, and ADR performed somewhat worse in terms of estimation efficiency. The converse, however, was not true. AUTO did not always perform as well as ADR when the adaptive scheme generated the data, and in two of the tests the differences were significant.

5) When the disturbances were generated by a second order autoregressive process, the results were slightly mixed. AUTO did marginally better than ADR for the larger samples and marginally worse for the smaller ones. Again, the differences were not significant. This, together with finding 4, indicates that ADR is a robust technique.

6) When the slope coefficients were subject to sequential change, ADR again performed remarkably well when the error was evaluated about the true value of the slope coefficients in period $T + 1$. Both the MSFE and MSEE are substantially smaller, indicating that ADR is robust in this situation as well.

7) A comparison of the estimation and forecasting results for equations A and B indicates that forecasting performance improves for both OLS and AUTO when an explanatory variable has a time trend. This is offset by an increase in the mean squared estimation error of $\beta_2$, the coefficient of that variable. Apparently, the time trend variable serves as a proxy for the intercept change.[14] When the intercept is subject to a constant change every period, the structure is equivalent to a constant intercept model with an omitted time trend. Thus, it is not surprising that OLS and AUTO forecast well in this situation since $x_2$ is highly correlated with time. This explains the exception noted in 2) above.

8) A comparison of the results for equation A with 35 and 20 observations indicates that sample size had a decided impact on performance. The ADR estimates had lower MSFE and MSEE, both absolutely and relative to OLS and AUTO when more observations were available. When intercept change was present, however, OLS performed significantly better on the small sample than on the large one, despite the fact that the MSEE for $\beta_1$ and $\beta_2$ were sometimes significantly higher and sel-

scheme and $\rho_1 u_{t-1} + \rho_2 u_{t-2}$ for the second order autoregressive scheme. The mean square errors of forecast differ from the mean square error about the conditional expectation by a constant; thus, drawing inference about MSFE about the conditional expectations rather than the actual realization is a valid procedure.

[13] Correlated sampling was used in this study to obtain the maximum information for a given sample size. We tested whether the difference in square errors between ADR and each of the other techniques was greater or equal to zero using the sign tests. $T$ tests were tried but because of the extreme observations the test was not appropriate even for 100 replications. Since the extreme observations were always in favor of ADR, namely large negative values, the median exceeds the mean making our test conservative. Summers (1965) uses such procedures.

Computer time constraints prevented us from increasing the sample size sufficiently so that a symptotic theory could be invoked and $t$ tests used.

[14] The detailed results revealed that the OLS estimates of $\beta_2$ had $t$ statistics that were two to five times those of the ADR estimates. The referee pointed out this result could have been anticipated given Theil's theorem on impact of omitted variable (1957).

dom significantly lower for the smaller sample. When there is no intercept change, the estimators show comparable improvements as the number of observations is increased.

9) The mean squared errors were principally the result of variance and not bias except for equation B with the constant change scheme. This is not surprising as, in all other cases, OLS provides unbiased forecasts and slope estimates. The nonlinear techniques AUTO and ADR had equally small bias, save for that one exception.

### III Time Trend Analysis

Frequently in econometric analysis equations are estimated with a time variable included. Sometimes it is argued that a trend variable is necessary to delineate growth effects from the permanent relationship being measured. It is also argued that time variables may arise naturally when the equation being studied is the solution to a dynamic system. Usually, however, the only apparent justification for the inclusion of a time trend is that its $t$ statistic is large in absolute value and there is some phenomenon in the system that is not accounted for by the other variables. We hypothesize that a significant time trend may be absorbing some of the effects of structural change. The purpose of the analysis reported here is to test the validity of this conjecture. The following two structures were used:

*Time Trend Present*

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 t + u_t \quad \text{for} \quad t = 1,2,\ldots$$

where $\beta_0 = 10$, $\beta_1 = 2$, $\beta_2 = 0.4$ and the $u_t$ were identically and independently distributed normal variates with mean 0 and variance 16.

*Adaptive Model*

$$\beta_{0,0} = 10.0,\ \beta_{0,t+1} \sim N(\beta_{0,t},8),\ \beta_1 = 2,\ \text{and}\ u_t \sim N(0,8).$$

In both cases the $x_{1t}$ were identically and independently distributed normal variates with mean 1 and variance 25. The time trend structure has a time trend and a constant intercept while the adaptive model has a random intercept change but no time trend. Both models were estimated under the assumption that a time trend was present and, alternatively, was absent by each of the three estimation procedures.

When the data were generated with a time trend included, the $t$ statistics of the time trend for OLS always exceeded 2. When the data were generated by the Adaptive Model, the $t$ statistic of the time trend for OLS exceeded 2 in 21 out of 25 trials and exceeded 5 in 14 out of 21 trials. OLS was clearly unable to differentiate between the two models. The AUTO and ADR techniques are more successful in differentiating between the two models. With ADR, however, the value of the $t$ statistics were far smaller and do not give the researcher a false sense of confidence in his estimates. The AUTO method falls somewhere between ADR and OLS.

In terms of estimating and forecasting efficiency, correct specification is crucial with OLS but not very important with ADR. Table 4 reveals that the MSFE increased only from 7.3 to 9.3 for the ADR estimates when time was incorrectly excluded while, with OLS, the MSFE increased from 4.4 to 56.9. When the adaptive model generated the data, the incorrect inclusion of time increased the MSFE from 12.9 to 17.8 for ADR and reduced it from 94.0 to 57.3 for OLS. Entries marked [b] indicate that the forecasting difference between ADR and both OLS and AUTO are statistically significant. Thus, when intercept variation is present, the performance of OLS is substantially improved if a time variable is arbitrarily included in the equation and, the time variable will most likely appear significant.

TABLE 4. — SUMMARY [a] OF TIME TREND ANALYSIS

| Estimation Procedures | Mean Square Forecasting Errors | | Mean Square Estimation Errors: $\beta_1$ | |
|---|---|---|---|---|
| | Time Trend Present | Adaptive | Time Trend Present | Adaptive |
| *Time Included* | | | | |
| OLS | 4.4 | 57.3 | .030 | .042 |
| AUTO | 5.3 | 28.5 | .028 | .024 |
| ADR | 7.3 | 17.8 [b] | .028 | .022 |
| *Time Excluded* | | | | |
| OLS | 56.9 | 94.0 | .037 | .130 |
| AUTO | 27.0 | 32.0 | .021 | .038 |
| ADR | 9.3 [b] | 12.9 [b] | .017 | .022 |

[a] The samples consisted of 100 trials.
[b] Indicates fraction of times ADR was better than OLS is significant at the 5 per cent level.

## IV Tests Using Friedman-Meiselman Models

The controlled experiments reported in the previous section indicate that adaptive regression is a robust and accurate technique. It is always of interest, however, to examine how an estimation technique performs on real data. In this section the adaptive regression technique is applied to the Friedman-Meiselman models. Friedman and Meiselman (1963) estimated the following model

$$C_t = \beta_0 + \beta_1 A_t + \beta_2 M_t$$

where $C_t$ is consumption, $A_t$ autonomous expenditures and $M_t$ the money supply. The purpose of their analysis was to test the relative stability of money velocity and the multiplier. The purpose of our analysis is not to attempt to answer this question but to test whether ADR forecasts better than OLS and AUTO. This particular model was selected because data were readily available for a long period of time, namely 1897–1958, and the structure involved no simultaneity.[15] Further, such a simple structure will surely be subject to structural change over time. Two additional versions of this model were estimated by Friedman and Meiselman. One, called the Keynesian model, constrained $\beta_2$, the effect of money, to be zero while the other called the quantity theory model, constrained $\beta_1$, the multiplier, to be zero.

In our tests we first took the data from 1897–1916 and estimated each of the versions of the model by OLS, AUTO and ADR. One and two-period forecasts were generated. The sample was then successively revised, new estimates obtained, and new forecasts generated for each of the years from 1917 to 1941, and 1947 to 1958. The sample was revised by adding one observation and deleting the earliest, so the sample size was always 20. This was done because we felt the assumptions of constant slope coefficients was not valid for longer periods. The results are summarized in table 5. Again, the performance of the ADR technique is dramatically superior to that of OLS and AUTO. For the years 1917–1937, OLS had slightly smaller MSFE than either ADR or AUTO with the combined model. In all other instances both OLS and AUTO did significantly worse than ADR for both one- and two-period forecasts.

[15] Friedman and Meiselman's definitions were used. Whether they are the appropriate or best definitions, we have no comment.

TABLE 5. — SUMMARY OF RESULTS FOR F-M MODELS

| | Forecast Years 1917–1937 | | | Forecast Years 1938–1958 | | |
|---|---|---|---|---|---|---|
| | OLS | AUTO | ADR | OLS | AUTO | ADR |
| *Combined Model* | | | | | | |
| One-Period Forecast | | | | | | |
| MFE | −1 | −1 | −1 | 9 | 4 | 4 |
| MSFE | 10 | 25 | 11 | 190 | 82 | 51 |
| Two-Period Forecast | | | | | | |
| MFE | −2 | −1 | −1 | 13 | 8 | 8 |
| MSFE | 14 | 20 | 17 | 342 | 164 | 143 |
| *Keynesian Model* | | | | | | |
| One-Period Forecast | | | | | | |
| MFE | 16 | −42 | 2 | 31 | −4 | 14 |
| MSFE | 374 | 2079 | 27 | 1642 | 1321 | 614 |
| Two-Period Forecast | | | | | | |
| MFE | 19 | −39 | 4 | 40 | 9 | 25 |
| MSFE | 489 | 1989 | 83 | 2432 | 1603 | 1301 |
| *Quantity Model* | | | | | | |
| One-Period Forecast | | | | | | |
| MFE | −3 | −1 | −1 | 7 | 3 | 2 |
| MSFE | 25 | 30 | 12 | 199 | 265 | 106 |
| Two-Period Forecast | | | | | | |
| MFE | −4 | −1 | −2 | 10 | 5 | 4 |
| MSFE | 30 | 41 | 19 | 365 | 332 | 239 |

## V Summary and Conclusions

Problems of structural change and misspecification, common in empirical research, are difficult, if not impossible, to deal with directly. The adaptive regression model, which assumes explicitly that the equation may be subject to permanent structural change over time, appears to be a valuable technique for dealing with such problems. Monte Carlo tests indicate that in terms of both forecasting and estimation efficiency adaptive regression is superior to ordinary least squares (with and without the autoregressive correction) when equations are subject to change over time. Adaptive regression is also remarkably robust when disturbances are generated by an autoregressive process. When the possibility of structural change over time is acknowledged, the significance of time trends in economic relations is questionable. Adaptive regression performs well even when time trends are incorrectly omitted. In summary, the tests reported in this paper indicate that because of its more general specification and robustness adaptive regression is a desirable alternative to conventional techniques.

REFERENCES


Brown, M., *On the Theory and Measurement of Technological Change* (Cambridge: Cambridge University Press, 1966).

Cooley, T., and Prescott, E., "The Adaptive Regression Model," Forthcoming in *International Economic Review*, 1973.

Dhrymes, P., "Efficient Estimation of Distribution Lags with Auto-correlated Errors," *International Economic Review,* 10 (1969), 47–67.

Duesenberry, J., G. Fromm, L. R. Klein, and E. Kuh, *The Brookings Quarterly Econometric Model of the United States* (Chicago: Rand McNally, 1965).

Duffy, W. J., *Parameter Variation in a Post War Econometric Model*, Unpublished Ph.D. dissertation, University of Pittsburgh, 1969.

Friedman, M., and D. Meiselman, "The Relative Stability of Money Velocity and the Investment Multiplier in the United States, 1897–1958," in Commission on Money and Credit, *Stabilization Policies* (Englewood Cliffs, N.J.: Prentice-Hall, 1963), 165–268.

Goldfield, S. M., "An Extension of the Monetary Sector," in Duesenberry et al. (eds.), *The Brookings Model: Some Further Results* (Chicago: Rand McNally, 1969), 319–362.

Howrey, P., "Structural Change and Postwar Economic Stability: An Econometric Test," this REVIEW (Feb. 1970), 18–25.

Klein, L., and M. Evans, *The Wharton Econometric Forecasting Model* (Philadelphia: The University of Pennsylvania, 1967).

Malinvaud, E., *Statistical Methods of Econometrics* (Amsterdam: North-Holland Publishing Company, 1966).

Quandt, R., "Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes," *Journal of the American Statistical Association* (1957), 873–880.

Rosenberg, B. M., *Varying Parameter Estimation*, unpublished Ph.D. thesis, Harvard University, 1968.

Summers, R., "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators," *Econometrica*, 33 (1965), 1–41.

Theil, H., "Specification Errors and the Estimation of Economic Relationship," *Review of the International Statistical Institute* (1957), 41–51.

Zellner, A., and G. Tiao, "Bayesian Analysis of the Regression Model with Auto Correlated Errors," *Journal of the American Statistical Association* (1965), 768–778.