1978

A comparison between unweighted and weighted regression for forecasting

Gilbert R. Valera
Iowa State University

A comparison between unweighted and weighted regression for forecasting

by

Gilbert R. Valera

A Thesis Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
MASTER OF SCIENCE

Major: Economics

Iowa State University
Ames, Iowa

1978

TABLE OF CONTENTS

CHAPTER I. INTRODUCTION

The use of weighted regression analysis is closely associated, from different points of view, with the building of an econometric model. The weights given to different observations are implicitly considered from the moment when the model builder establishes the purposes of constructing the model to the moment when it is decided to use a determined amount of data.

If the purpose of the model is to predict the magnitude and/or direction of sudden change of an endogenous variable, it would seem convenient to consider that some observations are more important than others. When the model is used for forecasting and it is required to compare and interpret mean square error statistics of an autoregressive model with a structural model, Howrey et al. (18, p. 376) suggest that power of forecast error comparisons might be increased by placing more weight on those periods during which the economy is undergoing unusual changes. If in the stage of choosing the variables which will be included in the model it is considered that these variables are measured with error, this will lead to the use of weighted regression. On the other hand, if the assumption that the variance is constant for all the observations is violated, the weighted regression is relevant. Even the use of less than all available data for estimation is one way of weighting the observations, to the nonused data in the procedure of estimation is assigned a weight of zero to each observation, and a weight of one to each observation used for estimation.

The method of Least Squares has the property that large deviations are treated with relatively greater attention (weight) than smaller ones; the weight assigned to these deviations would increase if the distribution had longer tails than the normal. Chow (5, p. 663) considers that a robust estimator which gives less weight to the large residual would be more acceptable for residuals which are nonnormally distributed.

One of the purposes of weighting is to give more attention to the measures of the independent variables, so that the best prediction is possible under the conditions of the relations among the explanatory variables themselves and between them and the dependent variable.

Thus, the processes of estimation and validation are not completely separate processes, and as is pointed out by Ladd (25, p. 10), if a criterion is sufficiently important to be used in validating a model, it is sufficiently important to be incorporated into the estimation procedure.

The objectives of this thesis are:

1. General

   a. To study the use of the weighted regression and the validation of econometric models.

   b. To consider the reasons for using weighted regression.

2. Specific

   a. To evaluate several econometric models in which the parameters have been estimated using a weighted regression procedure and an unweighted procedure.

   b. To apply different criteria of evaluation to the econometric models compared.

   c. To compare weighted regression procedure with unweighted regression for forecasting.

It is evident that no definitive conclusion can be reached on the analysis accomplished. However, it may be possible to indicate an example or model that can serve to the researcher to determine which type of data selection and which method of analysis he should use to utilize his data optimally.

Chapter II and III summarize literature on weighted regression and model validation. Chapter IV and V present a new method of weighted regression and model validation measures respectively. Chapter V presents empirical results. Chapter VI is a summary.

CHAPTER II. THEORY OF UNWEIGHTED AND WEIGHTED REGRESSION

Unweighted Regression

The model of regression tries to explain observed changes in a dependent variable (Y) as a consequence of changes in the independent variables $(X_1, X_2, \ldots, X_k)$. The functional relationship among the variables can be written as:

$$Y = f(X_1, X_2, \ldots X_k) + \varepsilon \tag{2.1}$$

where

ε: is a random variable called residual or error; this error is due to the fact that a perfect explanation of the dependent variable cannot be expected from the independent variables.

If the relationship among the variables is a linear function, then the model can be expressed as:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k + \varepsilon \tag{2.2}$$

If there are t observations on Y and each variable X_i; $i = 1, 2, \ldots, t$. The model becomes:

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i \tag{2.3}$$

where:

Y_i: is the i-th observation of the random dependent variable.

X_{ji}: is the i-th known value of j-th explanatory variable.

ε_i and X_j: are independent for all i and j.

X_1: is a dummy variable with value one.

β_j: is the j-th parameter and is to be estimated from the data.

ε_i: is a random variable with expectation zero, common variance $Var(\varepsilon_i) = \sigma^2$ for all i, and ε_i and ε_j are independent for all i and j.

In matrix notation the model is written as:

$$Y = X\beta + \varepsilon$$

where:

Y: is a t x 1 matrix.

X: is a t x k matrix and rank $k<t$.

β: is a k x 1 matrix.

ε: is a t x 1 matrix.

The assumption that the errors are statistically independent and have variance σ^2 can be expressed as:

$$E(\varepsilon\ \varepsilon') = \sigma^2 I_t \qquad (2.4)$$

where:

E( ) means expected value and I_t is a t x t identity matrix.

The property of equal variances is commonly referred as Homoscedasticity.

The Least Squares estimate of β is the estimate $\hat{\beta}$ which minimizes the residual sum of squares.

$$\sum_{i=1}^{t} e_i^2 = e'e = (Y-X\hat{\beta})'(Y-X\hat{\beta}) \quad (2.5)$$

Minimizing e'e with respect to $\hat{\beta}$ yields the Least Squares estimator, which is found to be:

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (2.6)$$

The variance of $\hat{\beta}$ is:

$$Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

An unbiased estimator of the residual variance is:

$$S^2 = \frac{e'e}{t-k}$$

Then

$$\hat{V}ar(\hat{\beta}) = S^2 (X'X)^{-1} \quad (2.7)$$

These estimators are both unbiased and consistent when the Least Squares assumptions are met.

The total sum of squares is:

$$SS(TOTAL) = Y'Y - t\overline{Y}^2$$

The total sum of squares due to the regression is:

$$SS(REG) = \hat{\beta}'X'Y - t\overline{Y}^2$$

The residual sum of squares is:

$$SS(RES) = SS(TOTAL) - SS(REG)$$

Thus, the mean square residual (S^2) can be also expressed as:

$$S^2 = \frac{SS(TOTAL) - SS(REG)}{t-k}$$

The coefficient of correlation is defined as the square root of:

$$R^2 = \frac{SS(REG)}{SS(TOTAL)} \qquad (2.8)$$

A value of R^2 close to unity means that the regression equation highly explains the variation of the dependent variable, an R^2 close to zero indicates that almost none of the variation on Y is explained by the independent variables.

The use of R^2 presents two major problems: First, it is assumed that the model is correctly specified, which is not necessarily true. Second, the addition of new explanatory variables increases the value of R^2. This can be resolved by using $\overline{R}^2$ instead of R^2. $\overline{R}^2$ also accounts for the number of degrees of freedom.

This statistic is defined as:

$$\bar{R}^2 = 1 - \frac{Var(e)}{Var(Y)} = R^2 - \frac{k-1}{t-k}(1-R^2) \tag{2.9}$$

The expected value of Y, given a fixed set of X_j's,

$X^* = (X_1^*, X_2^*, \ldots, X_k^*)$, is

$Y^* = X^*\hat{\beta}$

where:

Y^*: is a forecast of Y_t for the same period X^*.

$X^*\hat{\beta}$: is the best linear unbiased estimator (BLUE) for Y^*.

The estimate variance of Y^* is:

$$\hat{Var}(Y^*) = X^*(X'X)^{-1}X^{*\prime}S^2 \tag{2.10}$$

Let $\dot{Y}$ denote the actual value of Y for the period of forecasting.

Then the forecast error is given by:

$e_f = Y^* - \dot{Y}$

Under the assumption that the elements of ε^* are uncorrelated with ε, i.e.

$$VAR \begin{bmatrix} \varepsilon \\ \varepsilon^* \end{bmatrix} = \begin{bmatrix} \sigma^2 I & 0 \\ 0 & \sigma^2 I \end{bmatrix} = \sigma^2 I$$

The estimator of the variance of the forecast error is:

$$\hat{V}AR(e_f) = S^2[X^*(X'X)^{-1} X^{*\prime} + 1] \qquad (2.11)$$

There are two parts in this variance:

1. Sampling error of the LS coefficient estimator.
2. The random error e_f in the future observations.

To these two sources of forecasting error should be added two more:

3. The random nature of the additive error process guarantees that forecasts will deviate from true value, even if the model is specified correctly and its parameter values are known with certainty.
4. Error of specification in the model.

It can be shown that forecast error variance is minimized when all the new observations on the independent variables are equal to their mean values. The value of this is:

$$VAR(e_f) = \sigma^2[1 + \frac{1}{t}]$$

when t becomes sufficiently large, that is to say, the number of observations is very large, the variance of the forecast error approaches the variance of the error term.

The assumption that ε has a normal distribution with mean zero and matrix of covariance $\sigma^2 I_t$ implies that Y follows a t-variate normal distribution with mean vector $X\beta$ and covariance matrix $\sigma^2 I$. Thus $\hat{\beta}$ is normal with mean vector β and covariance matrix $\sigma^2 (X'X)^{-1}$; this allows us to derive confidence regions and tests of hypotheses.

These estimates accord with the maximum likelihood estimates.

Weighted Regression

Weighted regression has been developed by several authors according to different assumptions on the model; thus, it is possible to consider five cases:

1. The weighted regression as a consequence of the violation of the assumption that the residuals have a common variance. This is known as "Heteroscedasticity".
2. The weighted regression as a method of estimation which allows us to assign more importance to some observations than to others.
3. The weighted regression as a consequence of the variables in the regression equation being measured with error. That is called "Errors in Variable Model".
4. The weighted regression as a consequence of random coefficients.
5. The weighted regression as a consequence of estimating rational functions.

Let's be more explicit about each case in the next pages.

Case A: Heteroscedasticity

This is a common issue that appears in econometric books; for reference this can be seen in Theil (40, p. 244), Johnston (22, p. 214), Draper and Smith (7, p. 77), etc.

The assumption that the covariance matrix is:

$$E(\varepsilon\varepsilon') = \sigma^2\Omega$$

instead of:

$$E(\varepsilon\varepsilon') = \sigma^2 I$$

where:

Ω is a symmetric positive definite matrix of order t. This assumption leads to what is called the Generalized Least Squares Estimator (GLS) of β in the model

$$Y = X\beta + \varepsilon$$

For obtaining this estimator, it is necessary to transform the observation matrix [Y X] so the variance matrix is $\sigma^2 I$. Let T be the matrix transformation such that $|T| \neq 0$ and $\Omega^{-1} = T'T$. The transformation leads to:

$$TY = TX\beta + T\varepsilon$$

The Least Squares Estimator of β is the estimator b which minimizes the sum of squares

$$e'T'Te = e'\Omega^{-1}e = (TY-TXb)'(TY-TXb)$$

differentiating the last expression and equating it to zero, b is found to be

$$b = (X'T'TX)^{-1}X'T'TY \tag{2.12}$$

$$b = (X'\Omega^{-1}X)X'\Omega^{-1}Y \tag{2.13}$$

This estimator b of β, is the Best Linear Unbiased Estimator (BLUE).

The covariance matrix is:

$$VAR(b) = \sigma^2(X'\Omega^{-1}X)^{-1}$$

An unbiased estimator of σ^2 is:

$$S^2 = \frac{e'\Omega^{-1}e}{t-k}$$

From this, an unbiased estimator of the covariance matrix of b is:

$$VAR(b) = S^2(X'\Omega^{-1}X)^{-1} \tag{2.14}$$

If Ω is a diagonal matrix, say,

$$\Omega = diag(h_1, h_2, \ldots, h_t)$$

where:

$$h_i > 0$$

and which are, in general, different; then:

$V = \sigma^2\Omega$ is also diagonal,

V can be written as:

$$V = \text{diag}(\sigma^2 h_1,\ \sigma^2 h_2, \ldots, \sigma^2 h_t)$$

$$= \text{diag}(W_1, W_2, \ldots, W_t)$$

where:

$$W_i = h_1 \sigma^2{}_i$$

Thus, the errors ε_i are uncorrelated but have different variances. This situation is called Heteroscedasticity.

The transformation matrix T applied to the data [Y X] reduces the model to:

$$Y_i/\sqrt{W_i} = \sum_{j=1}^{k} \frac{\beta_j X_{ji}}{\sqrt{W_i}} + \frac{\varepsilon_i}{\sqrt{W_i}} \qquad (2.15)$$

for $i = 1,2,3...t$

In this way, the values of each observation are weighted inversely proportional to the standard deviation of the corresponding residuals. This is called Weighted Least Squares.

The normal equations are of the form:

$$\sum_i^t h_i X_{ij} Y_i = b_1 \sum_i^t h_i X_{ij} X_{i1} + \ldots,\ b_k \sum_i^t h_i X_{ij} X_{ik}$$

for $j = 1,2,3,\ldots,k$

Johnston (22, p. 212) shows how to use the GLS model for prediction.

If X^* is a vector of known values of the explanatory variables, the value of the dependent variable will be:

$$Y^* = X^*\beta + \varepsilon^*$$

where:

ε^*: is the unknown value of prediction disturbance.

It is assumed that:

$$E(\varepsilon^*) = 0$$

and

$$E(\varepsilon^{*2}) = \sigma^{2*}$$

If the residuals ε and ε^* are uncorrelated for the purpose of prediction it is necessary to use b, GLS estimator of β.

The covariance matrix of the prediction disturbance with the sample disturbance can be expressed as:

$$VAR \begin{bmatrix} \varepsilon \\ \varepsilon^* \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$$

Then the best unbiased prediction is:

$$Y^* = X^*b + Z_{21}Z_{11}^{-1}$$

where

$$b = (X'Z_{11}^{-1}X)^{-1}X'Z_{11}^{-1}Y \qquad (2.16)$$

and

$$e = Y - Xb$$

The prediction error has two parts when the Least Squares estimation is used, one error due to the sampling error, and another due to neglecting the future disturbances.

The points expressed about the hypothesis testing and confidence intervals are still applicable when the original observation matrix [Y X] is replaced by [TY TX], where T satisfies $T'T = \Omega^{-1}$.

b will be normally distributed with mean vector β and covariance matrix

$$\sigma^2 (X'\Omega^{-1}X)^{-1}. \tag{2.17}$$

The consequences of ignoring the different weights assigned to each variance and of estimating the parameters using OLS are two-fold. The estimates of the parameters are unbiased and consistent but have higher variances than the Least Squares estimators and the estimates of variances are biased; that is, as it is expressed by Pindick and Rubinfeld (30, p. 96):

> . . . ordinary least squares estimation places more weight on the observations which have large error variances than those with small error variances. The implicit weighting of ordinary least squares occurs because the sum of squared residuals associated with large variance terms is likely to be

substantially greater than the sum of squared residuals associated with low variance errors. . . .

Case B

The weighted regression can be considered as a method of estimation which allows us to assign more importance to some observations than to others. Commonly in economic series, some observations have a behavior quite different from the rest of the other observations, but omitting these observations because they do not follow the behavior of the larger part of data, does not seem the most reasonable procedure.

Sometimes, these observations reveal or indicate a peculiar activity in economy; i.e., the economic changes which were present in the Second World War or more recently the economic adjustments due to the increase of the oil prices.

A procedure frequently used is to estimate the regression equation and look at the residuals, and omit the observation with large residual; but this leads to discarding the standard errors and the confidence intervals constructed before. Fisher (10, p. 13) chose this procedure, preferring meaningful results of little precision to precise results of little or no meaning; but, if the model constructed is going to be used for forecasting, results with little or no precision are not very reliable.

Turning points is a criterion broadly used for validating econometric models; but, as it was suggested by Ladd, (25, p. 1) this would indicate that some observations are more important than others. Thus, the assignment of different weights to different observations would lead to a weighted regression. Also, if it is assumed that the residuals are uncorrelated (the matrix of weights is diagonal) this would be a case of Heteroscedasticity. Therefore, it is possible to question the validity of omitting observations because they are considered as "unusual".

This concern about the importance of some observations and, hence, the weights which should be assigned to each one, have led to alternative forms of estimation for the parameters in the regression equation. The robust regression, for example, as developed by Huber (19, p. 799) is one of them. Several works have shown, Andrews et al. (3, p. 89), and Chow (5, p. 663), that the method of Least Squares may be far from optimal if the distribution has large tails. Huber (20, p. 1041) indicates:

> . . . just a single grossly outlying observation may spoil the least squares estimates and moreover outliers are much harder to spot in the regression than in the simple location case. . . .

Based on this, Huber suggests minimizing:

$$\sum_{i=1}^{t}[f(Y_i - \sum_{j=1}^{k} X_{ij}\beta_j)] = \sum_{i}^{t} f(\varepsilon_i) \qquad (2.18)$$

where:

$$f(\varepsilon_i) = \begin{cases} 1/2\ \varepsilon_i^{\,2} & \text{for } |\varepsilon_i| < m \\ m|\varepsilon_i| - 1/2\ m^2 & \text{for } |\varepsilon_i| \geq m \end{cases} \tag{2.19}$$

m being a predetermined constant, if $m = \infty$, the estimation procedure is reduced to OLS. This is obviously a method of weighting the residuals.

The normal equations have the form:

$$\sum_{i=1}^{t} h\left(\frac{Y_i - \Sigma X_{ij}\beta_j}{\sigma}\right) X_{ij} = 0 \tag{2.20}$$

where

$$d = h\left(\frac{\varepsilon_i}{\sigma}\right) = \max[-m, \min(m, \varepsilon_i)]$$

The residual variance is found to be:

$$\frac{1}{t-k} \sum_{i}^{t} h\left[\frac{\varepsilon_i}{\sigma}\right]^2$$

Since the method of OLS gives more weight to the large fluctuations and bends the fitted regression into them to the disadvantage of the smaller fluctuations, it has been suggested to minimize the sum of absolute errors, the reason being that large deviations are not compensated disproportionately at the expense of smaller ones. This kind of estimator is defined as L_p estimator, which minimizes:

$[\Sigma|\varepsilon_i|^p]^{1/p}$

when:

p = 1, the estimator is called Least Absolute Residual (LAR).

p = 2, the estimator is OLS, thus the LAR estimator minimizes:

$$S = \sum_{i=1}^{t} |\varepsilon_i| = \Sigma[\frac{\varepsilon_i^2}{|\varepsilon_i|}] = \Sigma W_i \varepsilon_i^2 \qquad (2.21)$$

where:

$W_i = 1/|\varepsilon_i|$

Thus, the extreme deviations are almost ignored; this is clearly a weighted Least Squares problem. The solution to this problem can be faced as an iterative process using the reciprocal of the absolute values of the residual obtained by OLS as initial estimates of W's and then, to minimize:

$\sum_{i}^{t} W_i \varepsilon_i^2$, and repeat the procedure.

In a second estimator, the weights are defined as:

$$W_i = \begin{cases} [1-(\frac{Z_i}{K_1})^2]^2 & \text{if } |Z_i| < K_1 \\ 0 & \text{otherwise} \end{cases}$$

where:

$$z_i = \frac{\varepsilon_i}{K_2}$$

$K_1 = 6$ and $K_2 = m/.6745$, where m is a median of absolute values of the residuals.

These two methods of solution were suggested by Tukey (42) and used by Fair (9). Another solution, utilized by Fair, is to use OLS for small residuals and LAR for large ones. Once more the starting points are the residuals obtained by applying OLS.

A problem, which appears when this kind of estimation is used, is the construction of confidence intervals and hypothesis testing, since it is necessary to assume a distribution of the errors different from normal, the difficulty of constructing sample distribution of these estimators is present. If it is assumed that the absolute values of the residuals are distributed according to a double exponential distribution, then it is possible to obtain a maximum likelihood estimator which is equal to a LAR estimator; hence, these estimators will have all of the properties of ML estimators. Thus, the difficulty of constructing confidence intervals and tests of hypothesis is eliminated.

Several authors have proposed different distributions which all seem plausible thus the question about which distribution should be assumed remains.

Case C: Errors in variables model

The weighted regression can be considered as a consequence of the fact that the variables in the regression equation are measured with error. The use of a simple regression model, instead of a multiple regression one, will help to develop this idea. The exposition of Errors in Variable in a more general form can be seen in Fuller (14), or Zellner (44).

The model, in vector notation, for a sample of size t is:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i \\ X_i &= x_i + \mu_i \\ Y_i &= y_i + \varepsilon_i \end{aligned} \tag{2.22}$$

where

x_i and y_i: denote true values.

X_i: denotes observed values on x_i.

μ_i: represents the measurement errors in the X_i variables.

ε_i: represents the measurement error in Y_i.

β_0, β_1, y_i's and x_i's are unknown.

The assumptions for this model are:

1. x_i is fixed, that is, there is a true relation, with no error in the equation; this kind of model is called "functional".

2. measurement errors μ_i are distributed as independent normal with:

$E(\mu_i) = 0$

$Var(\mu_i) = \sigma_u^2$

μ_i: are independent of x_i and ε_i,

3. ε_i are independent normal $(0, \sigma_\varepsilon^2)$.

The likelihood function for the parameters β_0, β_1, σ_u^2, σ_ε^2 and x_i is given by:

$$L = A + \frac{1}{\sigma_u^t \sigma_\varepsilon^t} \exp[- \frac{1}{2\sigma_u^2} (X-x)'(X-x)$$

$$- \frac{1}{2\sigma_\varepsilon^2} (Y-q\beta_0-x\beta_1)'(Y-q\beta_0-x\beta_1)] \qquad (2.23)$$

where:

$\beta' = (\beta_0, \beta_1)$

$x' = (x_1, x_2, \ldots, x_t)$

$Y' (Y_1, Y_2, \ldots, Y_t)$

$X' = (X_1, X_2, \ldots, X_t)$

q: is a t x 1 column vector with each element equal to 1.

The maximization of the likelihood only requires minimization of the term in square brackets in L* = ln L. One more time the weighted regression is relevant, each component of the sum of squares is weighted inversely proportional to the size of the error variance. It is easy to make the analogy with the maximum likelihood estimators for OLS.

If $\mu_i \equiv 0$, this leads to OLS and it makes no difference whether ε_i is an error in the equation, is a measurement error, or both.

The determination of the estimators, see Zellner (44, p. 120, leads to:

$$\beta_1 = \frac{\sigma_\varepsilon^2}{\sigma_\mu^2}$$

A problem arises because, with each additional observation on (Y_i, X_i), it is necessary to estimate one additional parameter x_i. That is to say, the number of parameters estimated increases with the sample size. Then, one more assumption is required, namely, that the ratio of the error variances is known a priori.

$$\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\mu^2} \text{ is known,}$$

the likelihood function is:

$$L = A \frac{1}{\sigma_\varepsilon^{2t}} \exp - \{\frac{1}{2\sigma_\varepsilon^2}\{\frac{1}{2\sigma_\varepsilon^2}[\lambda(X-x)'(X-x)$$

$$+ (Y-q\beta_0-x\beta_1)'(Y-q\beta_0-x\beta_1)]\} \quad (2.24)$$

On differentiating L* with respect to the unknown parameters, and setting these derivates equal to zero, the following simultaneous equation system is obtained:

$$\frac{\partial L^*}{\partial \beta_0} = \frac{1}{\sigma_\varepsilon^2} (Y-q\beta_0-x\beta_1)q = 0$$

$$\frac{\partial L^*}{\partial \beta_1} = \frac{1}{\sigma_\varepsilon^2} (Y-q\beta_0-x\beta_1)x = 0$$

$$\frac{\partial L^*}{\partial x} = \frac{1}{\sigma_\varepsilon^2}[\lambda(X-x) + (Y-q\beta_0-x\beta_1)] = 0$$

$$\frac{\partial L^*}{\partial \sigma_\varepsilon} = - \frac{2t}{\sigma_\varepsilon} + \frac{1}{\sigma_\varepsilon^3}[\lambda(X-x)'(X-x) + (Y-q\beta_0-x\beta_1)'$$

$$(Y-q\beta_0-x\beta_1)] = 0$$

These partial derivates of L* contain variances as weights. Solving these equations, it yields:

$$\hat{\beta}_1 S_{XY} - \hat{\beta}_1 (S_Y^2-\lambda S_X^2) - \lambda S_{XY} = 0$$

From this:

$$\hat{\beta}_1 = \frac{S_Y^2-\lambda S_X^2 + [(S_Y^2-\lambda S_X^2+4\lambda S_{XY}^2)]^{1/2}}{2S_{XY}} \quad (2.25)$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}$$

where:

S_Y^2, S_X^2, and S_{XY} are the sample moments.

The choice of the positive sign of the square root is justified since this leads to a maximum of likelihood function. The knowledge of σ_ε^2 and σ_μ^2 permits the construction of approximate confidence intervals and hypothesis testing. This is not possible when only λ is known.

If x_i is considered a random variable the model becomes what is called a structural model; the estimators of β_0 and β_1 for this kind of model have the same characteristics as for the functional models.

What is relevant to this analysis is that the errors in variables model can be thought of as a weighted regression model where the weights are inversely proportional to size of the error variance.

Case D: Regression models with random coefficients

Many times, it is convenient to consider models with random coefficients; this is justified in studies on cross section data in which the parameters are not homogeneous among different cross section units.

The Engel curves is a typical case, Fisk (11, p. 266) points out that:

> Indeed, insofar as a regression model adequately describes the observed heteroscedasticity, information on the variability of the regression coefficient should be as important to the applied

economist as information on the mean values of those regression coefficients. . . .

On the other hand, in the estimation of an econometric model obtained from a problem of maximization or minimization, the involved variables determine the parameters of the model: hence, changes in these variables affect the parameters and this will lead to an econometric model with random coefficients. This aspect is truly important when the involved variables are policy variables and it is necessary to analyze the effects of economic policies. Nelder (29, p. 303) developed this model with random coefficients.

Let the model be:

$$Y_i = \beta_{0i} + \beta_{1i}X_i$$

where:

β_{0i} and β_{1i} are normal randomly distributed with means β_0 and β_1 and variance matrix:

$$Var(\beta_0, \beta_1) = \begin{bmatrix} \sigma^2_{\beta_0} & \sigma_{\beta_0\beta_1} \\ \sigma_{\beta_1\beta_0} & \sigma_{\beta_1}^2 \end{bmatrix}$$

Thus, the distribution of Y_i given X_i is also normally distributed with mean:

$$E(Y_i) = \beta_0 + \beta_1 X_i$$

and variance:

$$\mathrm{Var}(Y_i) = \sigma^2_{\beta_0} + 2\sigma_{\beta_0\beta_1} X_i + \sigma^2_{\beta_1} X_i^{\,2}$$

The logarithm of the likelihood function, L* is the proportional to

$$-\Sigma[\ln(\sigma^2_{\beta_0} + 2\sigma_{\beta_0\beta_1} X_i + \sigma^2_{\beta_1} X_i)]$$

$$- \Sigma[\frac{(Y_i-\beta_0-\beta_1X_i)^2}{\sigma^2_{\beta_0}+2\sigma_{\beta_0\beta_1}X_i+X_i^{\,2}\sigma^2_{\beta_1}}] + A \qquad (2.26)$$

where A is a constant.

The normal equations are obtained by differentiating L* with respect to β_0, β_1, $\sigma^2_{\beta_0}$ and $\sigma\beta_1\beta_0$; this procedure leads to:

$$\Sigma W_i(Y_i-\beta_0-\beta_1X_i) = \Sigma W_i e_i = 0 \qquad (2.27)$$

where:

$$e_i = (Y_i-\beta_0-\beta_1X_i)$$

$$\Sigma W_i X_i e_i = 0$$

$$\Sigma W_i(1-e_i^2W_i) = 0$$

$$\Sigma W_i X_i(1-e_i^2W_i) = 0$$

$$\Sigma W_i X_i^{\,2}(1-e_i^2W_i) = 0$$

where:

$$W_i = \frac{1}{\sigma^2_{\beta_0} + 2\sigma_{\beta_0\beta_1} + \sigma^2_{\beta_1} X_i^{\,2}} \qquad (2.28)$$

The knowledge of $\sigma^2_{\beta_0}$, $\sigma_{\beta_0\beta_1}$ and $\sigma^2_{\beta_1}$ would imply a problem of weighted regression.

The transformation of the variances by doing:

$$\sigma^2_{\beta_0} = \lambda \sin^2 \theta$$

$$\sigma^2_{\beta_1} = \lambda \cos^2 \theta$$

$$\sigma_{\beta_0\beta_1} = \lambda \sin \theta \cos \theta \sin \phi$$

$$\lambda > 0$$

$$0 \leq \theta \leq \pi/2$$

$$-\pi/2 \leq \phi \leq \pi/2$$

guarantees that the variance matrix be positive (semi)-definite

$\rho_{\beta_0\beta_1} = \sin \phi$, and λ is a scaling factor.

Using this transformation, L* may be expressed as:

$$\sum_{i}^{t} \ln(W_i^*/\lambda) - \sum_{i}^{t} \frac{W_i^*(Y_i-\beta_0-\beta_1 X_i)^2}{\lambda} + A$$

where:

$$W_i^* = [\sin^2\theta + 2X_i \sin\theta \cos\theta \sin\phi + X_i^2 \cos^2\theta]^{-1}$$

a maximum occurs when:

$$\lambda = \frac{\sum_i^t W_i^* e_i^2}{t}$$

This result reduces the estimation procedure to maximize:

$$\sum_i^t \ln W_i^* - t \ln[\sum_i^t W_i^*(Y_i - \beta_0 - \beta_1 X_i)^2] \qquad (2.29)$$

which can be considered a weighted least square problem.

The method for maximizing the above expression is to start with initial estimates of θ and ϕ for obtaining estimats for β_0, β_1 and ε_i, and a searching procedure is used for reaching a maximum.

Similar models have been developed by other authors, but commonly they present substantial computational difficulties.

This kind of model is very important because its characteristics are representative of several economic situations.

Case E: Estimation of parameters in a rational function

This problem is considered by Turner et al. (43, p. 120).

Although their aims are toward the solution of practical biological situations, rational functions can be used for describing any asymptotic process in general.

The model can be expressed as

$$Y = \frac{F(X)}{G(X)} + \varepsilon \tag{2.30}$$

Where:

F(X) is a polynomial of degree n

G(X) is a polynomial of degree k

ε: is normally distributed $(0, \sigma_I^{\ 2})$

$$E(\varepsilon_i\ \varepsilon_j) = \begin{cases} \sigma^2 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

It is assumed that:

$$F(X) = \eta_0 + \eta_1 X_1 + \eta_2 X_2^{\ 2} + \ldots + \eta_n X_n^{\ n}$$

$$G(X) = 1 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_k X^k$$

Thus the expression for Y becomes:

$$Y = F(X) - [\beta_1 XY + \beta_2 X^2 Y + \ldots + \beta_k X^k Y] + G(X)\varepsilon$$

If $[G(X)]^{-2}$ was known, the last expression would lead to the estimation of a weighted regression equation.

In matrix form the model can be written as:

$$X = \begin{bmatrix} 1 & X_1 \ldots X_1^n & X_1Y_1 \ldots X_1^kY_1 \\ 1 & X_2 \ldots X_2^n & X_2Y_2 \ldots X_2^kY_2 \\ \vdots & & \\ 1 & X_t \ldots X_t^n & X_tY_t \ldots X_t^kY_t \end{bmatrix}$$

$$\beta' = (\eta_0, \eta_1 \ldots, \eta_n, \beta_0 \beta_1 \ldots \beta_k)$$

$$Y' = (Y_1, Y_2, \ldots Y_t)$$

The matrix weight W

$$W = \text{diag}(G_1^{-2}, G_2^{-2}, \ldots G_t^{-2}) \qquad (2.31)$$

The estimation procedure can be accomplished by using an iterative process. Preliminary estimates of β's are utilized for obtaining provisional weights for computing improved estimator on η's and β's, new weights are found and the process is repeated until stable values are found. After the last iteration the estimator vector of the parameters is given by:

$$\hat{\beta} = (X'WX)^{-1}X'WY$$

$$VAR(\hat{\beta}) = (X'WX)^{-1}S^2 \qquad (2.32)$$

where:

$$S^2 = \frac{Y'WY - \hat{\beta}'X'WY}{t-k-n-1}$$

Asymptotic confidence limits for each coefficient can be obtained by using:

$$\hat{\eta}_i \pm t_c \sqrt{C_{ii} S^2} \qquad i = 0,1,\ldots,n$$

$$\hat{\beta}_i \pm t_c \sqrt{C_{ii} S^2} \qquad i = 1,2,\ldots,k$$

where:

C_{ii} is the i-th element in the diagonal of (X'WX) and

t_c is the critical value for a Student's t with (t-k-n-1) degrees of freedom.

CHAPTER III. MODEL VALIDATION

There are three major methodological positions concerning the problem of verification in economics. These three positions are: Rationalism, Empiricism, and Positivism.

From the point of view of Rationalism the problem of verification is viewed as a problem of searching for a set of basic assumptions underlying the behavior of the system of interest.

For Empiricism and in particular for T. W. Hutchison (21) the validity of a model depends on the validity of the assumptions on which the model is based.

Milton Friedman (12), as a representative of the Positivism position, argues that the validity of a model should be judged by its ability to predict the behavior of the variables.

Since econometric models are constructed for particular uses and specific tests are designed for their validation, Dhrymes et al. (6, p. 310) and Shapiro (35, p. 253) have pointed out that model validation is problem-dependent or decision-dependent.

Thus the model can be valid for one purpose and not for another.

The evaluation of an economic model involves two stages:

1. The stage referred to as the construction of the model. Several activities are considered in this stage:
   a. Formulation of hypothesis about the structure which the model is supposed to represent.
   b. Model selection: specification of variables and functional forms, and the method of estimation.
   c. Hypothesis test which the model will be based on.
2. This stage includes the evaluation of the econometric model as a whole. This second stage will be used in this thesis for evaluating the models chosen. Henceforth, the word evaluation will refer to the steps subsequent to construction of the model.

Parametric Evaluation

The parametric evaluation of an econometric model is based on statistical tests which are linked to the stochastic specification assumed by the builder of the econometric model. This parametric evaluation concerns both stages mentioned above. Statistical tests have been designed for testing hypotheses on model selection, optimal parameter estimation, forecasting evaluation and structural stability of the model. Ramsey's test (32, p. 351) may be used for testing for the presence of specification errors by comparing the distribution of the residuals under the alternative hypothesis. The estimation of parameters can be improved if it is known that the parameters are subject to constraints. Aitchison and Silvey (2) developed a test involving a Lagrange multiplier approach to the testing of a set of restrictions on the parameters being estimated.

The availability of a small data set, not used in estimation, raises the possibility of checking the model after it has been estimated.

Chow (4, p. 591) developed a test which may lead to the inference of structural change either because the coefficient vector β in the model $Y = X\beta + \varepsilon$, is different for each of the two sampled periods under investigation, or because the variance of ε has changed, or even for both of these causes. Jorgenson et al. (23, p. 216), based on this Chow's test, develop two test statistics, one based on predictive performance and the other on structural change. The data which lies outside the sample period is utilized for generating forecasts of the dependent variable which can then be compared with the actual values of the dependent variables; the test indicates whether the additional observations are from the same regression as those observations used in the sample period.

This test of predictive performance is used for comparing the error of prediction with errors of the fitted function; this test is based on a prediction interval for the mean of m additional observations. This test is useful in detecting error of specification.

The test of structural change is more powerful from the statistical point of view than the predictive test. This difference arises from the fact that the predictive test depends on the difference in the parameters for the period of fit and the period of prediction, and the difference of the residual variance; hence there are two components and the first one can hide the second one. In short, because the differences in the parameters are associated with errors of specification, the test for structural change is the best for detecting this kind of error.

This predictive test can be extended to a forecast of m new observations on each of several endogenous variables using the reduced from of a linear simultaneous equation model; however the assumptions made are quite restrictive, i.e.; the equations of the system must be just identified.

If the data set is considered to be quite big, it is possible to consider re-estimating the model. When this is the situation, the Chow's test gives the best results. The tests mentioned in this section are designed for testing hypothesis prior to "release" of the model, but this stage of the model building is not considered in this thesis. Ramsey's test, which would seem to be useful, assumes that one of the models being considered is the true model; this is a strong assumption which is not considered at all in this thesis. On the other hand, a specific objective of this thesis is to evaluate estimation procedures, hence the application of these tests is not relevant to this work.

Although the Chow's test can be used after the "release" of the model, the availability of a data set of new observations is also not taken into consideration in this thesis.

Nonparametric Evaluation

The evaluation process is defined to be nonparametric if it is not derived from stochastic assumption of the model.

There are several criteria based on this kind of evaluation:

Historical simulations

The primary criterion for judging the validity of a model is its power in explaining the observed data. Performing a historical simulation and examining how closely each estimated endogenous variable tracks its correspondent historical data series, may be the most useful criteria for evaluating a model.

When it is necessary to compare different types of functions, it is common to use a relative measure of fit (such as $\overline{R}^2$), but it is logical to think that the fit will be better in the sample period than outside the sample period, which indicates that this criterion should not be considered as the best one.

A sudden change in the historical data (turning point) is an important criterion for the model evaluation. If it is desirable that a model simulates the turning points, then this would seem to suggest that it would be convenient to consider that there exists some observations which are more important than others. This fact can be used for assigning different weights to different observations. This would lead to using weighted regression instead of classical least squares estimation.

The number of turning points in a time series can be used for testing the hypothesis that the series are random. Many turning points would indicate that the functions are not due to change alone.

This is not convenient for testing the hypothesis of linear trend series; the number of turning points is indifferent to the presence of a trend. A better way is to test the significance of the correlation coefficient or use the rank test (τ).

Nonstochastic A historical or ex-post simulation is called nonstochastic if the assumption is made that additive error terms are zero in each estimated equation in each observation period. The estimated coefficients are treated as if they are correct ones.

The application of nonstochastic simulation to nonlinear models yields results which are not consistent with the reduced form.

Stochastic   A stochastic simulation is done if, for each equation of the model, a probability distribution is assumed for the additive error term or for each estimated coefficient.

Since the coefficients are random variables and each equation has an additive error term associated with it, the stochastic simulation allows us to recognize the random character of the model.

The nonstochastical and stochastical simulations can be either static or dynamic.

Response to stimuli

The question which arises here is how the model responds to large changes in the exogenous variables or policy parameters.

These responses should be consistent with the economic theory and with empirical observations. Researchers have sometimes an idea of the range of variation of a specific coefficient. But the use to be made of this information depends on the preferences and experiences of the researcher and the actual needs for the user.

Using multipliers, it is possible to predict the effect of fiscal policy or deviations from any basic prediction path. It is easy to find the impact multipliers for a linear system but this is static value and does not show the accumulated effect if the change is sustained over many periods and it is assumed that only one exogenous variable changes. Part of these limitations can be relaxed by using dynamic multipliers.

Predictive ability

It has been expressed by Dhrymes et al. (6) and Shapiro (35, p. 255)

> . . . the evaluation of the predictive ability of the model is essentially a goodness of fit problem. . . .

This criterion has been considered powerful for validating a model.

The goodness of fit can be overstated when the economist considers a wide range of alternatives and selects the one that fits best.

Jorgenson et al. (23, p. 215) shows that the likelihood of achieving any predetermined level of goodness-of-fit can be made arbitrarily close to unity by expanding the range of alternatives considered, but makes clear that there exist two facts which are necessary to consider:

1. The research does not have many alternatives.
2. The test statistics for alternative specifications are not independently distributed, which violates one of the assumptions used.

These two facts make it difficult to achieve an optimal level of goodness-of-fit.

A forecast or prediction is generally defined as a statement concerning future events. It is possible to consider two kinds of forecasts:

1. Ex-post forecast in which the forecast period is such that the observations on both endogenous and exogenous variables are known with certainty.
2. Ex-ante forecast in which the explanatory variables may or may not be known with certainty. This sort of forecast predicts values of the endogenous variable beyond the estimation period.

A forecast is unconditional if all the values of the explanatory variables are known with certainty; otherwise the forecast will be conditional.

It is possible to consider two aspects in the evaluation of forecast performance:

1. Subjective: the evaluation is carried out by taking into consideration the capacity of the model to predict turning points, the size of the errors. It

would be desirable to detail the sources of deviation from realized values.

2. Objective: the criterion of forecast performance can be defined in terms of a loss function of the users of forecasts.

A main goal sought by researchers is to forecast turning points or predict the magnitude of change. Theil (39, p. 22) indicates two kinds of errors in predicting turning points:

1. A turning point is predicted but there is no actual turning point.
2. There is a turning point but it was not predicted before it happened.

Obviously, a test of a model's performance in predicting turning points is clearly important in overall appraisal. If a turning point is correctly predicted this indicates that the critical dynamic elements have been taken into account and the model is reliable when economic activity experiences changes in direction.

This kind of test can be applied with varying degrees of rigor. One case would be that all turning points must be correctly forecast, that would imply that the possibility of rejecting the model would be very high. An alternative criterion is that forecast shows a directional change in the neighborhood of the actual turning point. The latter criterion is less rigorous than the first one.

Stekler (37, p. 724) develops one hypothesis to explain why the turning point errors in the neighborhood of cyclical peak might have occurred.

Using the Bayesian approach, he assigns subjective possibilities to the likelihood of a cyclical turn. He estimates, based on information from economic indicators, the probabilities of occurrence of a signal from an indicator given a turn or given no turns; from this he obtains, using the Bayes' theorem, the probability of a turn given that a signal from an indicator has been received.

This analysis faces the problem of assigning probabilities to an event and the choice of the indicator.

3. Accuracy of the forecast.

Several measures can be used for comparing the predictive values of the endogenous variables with their actual values.

1. Mean Square Error

This method is defined by Mincer and Zarnowitz (27, p. 7) as:

$$MSE = E(A_t - F_t)^2 \qquad (3.1)$$

where:

A_t: is the actual value.

F_t: is the forecasted value.

It is based on a quadratic loss function, namely,

$$L(A-F_t) = (A_t-F_t)^2$$

which represents the loss resulting from the forecast F_t when A_t is a true value. This loss function leads to the estimator with a minimum second order sampling moment around the true value. MSE can be also expressed as:

$$MSE = \sum_{t=1}^{T} \frac{(A_t-F_t)^2}{T} \qquad (3.2)$$

where:

T: is the number of observations.

The MSE is a measure of dispersion around the line of perfect forecasts (LPF). This LPF is a 45° line through the origin, which is used for analyzing absolute forecast accuracy. Plotting the actual and forecast values in a scatter diagram, it is possible to fit a straight line:

$$A_t = \alpha + \beta F_t$$

This regression line should coincide with the LPF. Thus, MSE is equal to zero if all points lie on LPF. The forecast is unbiased if $E(F) = E(A)$; the bias is defined as:

$$E(\mu) = E(A) - E(F)$$

The larger the deviation of the slope of the regression line from unity, the less efficient the forecast.

The sample MSE can be decomposed as follows:

$$MSE = S^2_{(A-F)} + (\bar{A}-\bar{F})^2 \tag{3.3}$$

where:

$\bar{A}$ and $\bar{F}$ are mean values.

$S^2_{(A-F)}$ is the sample variance of the prediction errors.

The MSE can be also expressed as:

$$MSE = (\bar{A}-\bar{F})^2 + (1-R^2)S^2_A + (1-\hat{\beta})S^2_F \tag{3.4}$$

where:

S^2_A and S^2_F are sample variances.

$\hat{\beta}$ is the slope of the regression line.

R^2 is the coefficient of determination in the regression of A on F.

$(\bar{A}-\bar{F})^2$ is called mean component (MC).

$(1-R^2)S^2_A$ is the residual component (MR).

$(1-\hat{\beta})S^2_F$ is the slope component (MS).

then:

$$MSE = MC + MS + MR \tag{3.5}$$

If the forecast is unbiased then MC = 0.

If the forecast is efficient then MS = 0.

If the forecast is both unbiased and efficient then $MC = MS = 0$ and $MSE = MR$.

MC and MS can be interpreted as the proportion of error resulting from systematic tendencies of the forecast system.

Since the MSE and its components are calculated using a sample, they are subject to sampling variation, even if the estimates are unbiased and efficient in the population. It is possible to test unbiasedness and/or efficiency by testing the null hypothesis

$$H_o \alpha: = 0 \quad \beta = 1.$$

A problem arises when the errors are measured in terms of levels, they can combine overtime; hence they fail to show the true magnitude of the error. A common solution to this problem is to express the errors in terms of changes instead of levels.

If errors are going to be expressed in terms of changes it is necessary to choose the data that will be used.

There is no ambiguity if the model predicts changes; the choice of the data is evident. But, if the model predicts levels of economic variables, there are two ways to obtain the predicted changes data.

1. Successive differences of the predicted levels

$$\Delta F_t = F_t - F_{t-1}$$

2. Differences between the predicted values for a period and the actual values of a previous period:

$$\Delta F_t = F_t - A_{t-1}$$

The model's ex-post forecast error with this approximation is the difference between the actual and the predicted level.

Hence, the mean square error (3.2) would be identical, but the decomposition would be different; the variance and residual components would not be the same since the regression would be

$$(A_t - A_{t-1}) \quad \text{on} \quad (F_t - A_{t-1})$$

instead of A_t on F_t.

A problem arises when the predicted change $(F_t - \dot{A}_{t-1})$ is compared with the realized change $(A_t - A_{t-1})$, where $\dot{A}_{t-1}$ was not fully known at the time that the forecast was made. Only in the case where $\dot{A}_t$ is exactly known is the accuracy of the MSE for changes almost identical with the accuracy for levels.

Mincer and Zarnowitz (27) developed a measure of relative accuracy. Their index of forecasting quality is the ratio of the mean square error of forecast to the mean square error of extrapolation (ME_x). The use of ME_x is justified since it is a relatively simple, quick and accessible alternative.

Denoting the relative mean square error by RM, then:

$$RM = \frac{MSE_F}{ME_x} \qquad (3.6)$$

if $0<RM<1$ the forecasts are relatively superior to extrapolation.

Several objections can be made to the use of this technique. The use of a quadratic loss function, such as has been defined in expression (3.1), is justified in part because of its tractability, but its use implies that either kind of error (under and overestimations) is evaluated equally and high weights are assigned to the extreme errors.

Although R^2 is used as supplementary measure it is not a reliable guide for accuracy because it merely represents error explained by a linear adjustment of the forecast series. The MSE only evaluates forecasts in terms of systematic errors.

Granger and Newbold (16, p. 281) developed a test for equality of expected square forecast errors, when there are two or more sets of forecasts of the same quantity. This test is very important for comparing two competing forecasting procedures and it is based on the usual test for zero correlation.

$$r = \frac{\sum_{t=1}^{T}(e_t^{(1)} + e_t^{(2)})(e_t^{(1)} - e_t^{(2)})}{[\sum_{t=1}^{T}(e_t^{(1)} + e_t^{(2)})^2 \sum_{t=1}^{T}(e_t^{(1)} - e_t^{(2)})^2]^{1/2}} \qquad (3.7)$$

where: $e_t^{(i)}$ is one-step-ahead forecast error of the same quantity using the i-th (i = 1,2) prodedure. These errors are assumed to have a normal distribution with means zero, variances σ_{e1}^2 and σ_{e2}^2 and the correlation coefficient ρ.

The necessary and sufficient conditions of equality of the two expected square errors from the two forecasting methods are that the sample correlation coefficient between the two sets of errors forecasts (r), given by the formula (3.7), be zero. Rejecting the null hypothesis would indicate that one procedure performed significantly better than the other.

Granger and Newbold (16, p. 286) define a statistic for judging forecast performance. They insist on the use of predicted and actual changes instead of levels.

Let F_t be a predictor of A_t, and e_t be the forecast error,

$$A_t = F_t + e_t \qquad (3.8)$$

If the forecast and error series are uncorrelated, it is possible to define PM as the ratio of error variance to variances of the series to be forecast, so that

$$PM = \frac{\sigma_e^2}{\sigma_A^2} \tag{3.9}$$

PM = 0 if F is a perfect forecast.

PM = 1 if F is the mean of A for all t.

If F_t is an optimal forecast, say

$E(A_t) = E(F_t)$ and $Var(A_t) = Var(F_t) + Var(e_t)$ and is based on the particular information set, PM can be considered as a measure of the predictability of a time series.

If F and e are uncorrelated, $PM = 1 - \rho^2$. (3.10)

where

ρ: is the correlation coefficient between the actual and forecast values.

The probability that the forecast and actual series will have the same sign if they have zero mean, is given by:

$$P = 1/2 + (1/\pi)\text{arc sin } \rho = 1/2 + (1/\pi)\text{arc cos PM} \tag{3.11}$$

It is based on the assumption that the actual and forecast series are distributed as normal bivariate with correlation ρ.

2. U-Statistic

Theil (39, p. 28) defines a measure that utilizes information about the absolute discrepancy between the predicted and actual changes. U is defined as follows:

$$U = \frac{\sqrt{\sum_{t=1}^{T} \frac{(P_t - A_t)^2}{T}}}{\sqrt{\sum_{t=1}^{T} \frac{A_t^2}{T}} + \sqrt{\sum_{t=1}^{T} \frac{P_t^2}{T}}} \qquad (3.12)$$

where:

P_t: is the predictor change.

A_t: is the observed change.

T: is the number of observations.

U: is bounded: $0 < U < 1$

$$U = \begin{cases} 0 & \text{if } P_t = A_t \\ 1 & \text{if } P_t = -bA_t \text{ for } b > 0 \end{cases}$$

Theil (39, p. 28) modifies this statistic by signaling that in the U-statistics, the denominator depends on the forecast, hence the coefficient is not uniquely determined by mean square of prediction. He suggests the use of the U-statistics defined by root square of:

$$U^2 = \frac{\sum_{t=1}^{T} (P_t - A_t)^2}{\sum_{t=1}^{T} A_t^2}$$

$$U = \sqrt{\frac{MSE}{\frac{\Sigma A_t^2}{T}}} \qquad (3.13)$$

It is called inequality coefficient

U: is not upper bounded

U: is equal to zero if $P_t = A_t$

U: is equal to one when the prediction procedure leads to the same MSE as naive no change extrapolation.

MSE can be decomposed as:

$$MSE = S^2_{(A-P)} + (\bar{A}-\bar{P})^2$$

$$= (\bar{A}-\bar{P})^2 + (S_P-S_A)^2 + 2(1-R)S_PS_A$$

$$= (S_P-RS_A)^2 + (1+R^2)S_A^2 + (\bar{A}-\bar{P})^2 \qquad (3.14)$$

where:

$(S_P-S_A)^2$ is the variance component.

$2(1-R)S_PS_A$ is the covariance compoennt.

(S_P-RS_A) is the regression component.

$(1-R^2)S_A^2$ is the disturbance component.

$(\bar{A}-\bar{P})^2$ is the bias component.

The variance component gives an idea about the influence of the variance of the actual values.

The disturbance component is the variance of the residuals of the regression of the observed values on the predicted values.

Dividing both sides of the two last equations by MSE yields two sets of proportions:

1. The first one: $1 = U^M + U^S + U^C$

2. The second: $1 = U^M + U^R + U^D$

where:

U^M, U^S and U^C: are the bias, variance and covariance proportions respectively.

U^R and U^D: are the regression and the disturbance proportions.

$U^M + U^S + U^C = U^M + U^R + U^D$

Theil (39, p. 29) indicates that:

a. The term $(\bar{A}-\bar{P})^2 = 0$, if and only if the average predicted changes are equal to the average observed changes.

b. The term $(S_P-S_A) = 0$, if and only if $S_P = S_A$, and

c. The last term $2(1-R)S_PS_A = 0$, if and only if $R = 1$.

But Jorgenson, Hunter and Nadiri (23, p. 219), based on the predictive testing, point out that bias component has expected value different from zero. They say:

> Thus, even for the unique, minimum variance, unbiased, linear predictor, there is no reason for the bias component to be zero. . . .

There is also no reason for the variance component to be zero. They conclude that this criterion for evaluating a predictor is of no assistance in the predictive testing of an econometric model.

3. Mean Absolute Error:

The mean absolute error (MAE) is defined as:

$$MAE \sum_{t=1}^{T} \frac{|Y_{it}-Y_{Pit}|}{T} \tag{3.15}$$

where:

Y_{it}: is the actual value of the variable for the period t.

Y_{Pit}: is the predicted value of the variable for the same period.

This statistic is easy to compute and does not penalize the extreme errors highly. The MAE is highly correlated with the square root of MSE; this correlation is around .80 for normal and rectangular distribution.

The correlation coefficient provides a guide to accuracy of forecast. Small or negative values of this coefficient diminish the confidence in forecast even when the mean error is small.

The use of MAE is justified when the variable exhibits a steady trend: hence it is interesting to know how far above or below the actual trend line is the predicted series. In this way, the problem of positive and negative errors cancelling each other is avoided. If the errors are expressed in terms of changes, the mean absolute error can be expressed as:

$$\Delta MAE = \sum_{t=1}^{T} \frac{|(Y_{it}-Y_{it-1}) - (Y_{Pit}-Y_{Pit-1})|}{T} \qquad (3.16)$$

where:

Y_{it-1}: is the observed value of the variables for the period (t-1).

Y_{Pit-1}: is the predicted value of the variable for the period (t-1).

MAE and ΔMAE are the same for the one-quarter ahead forecasts.

4. Other statistics:

Other commonly used measures are based on percentages or percentages changes. The mean percent error (MPE) is defined as:

$$MPE = 1/T \sum_{t=1}^{T} \frac{(Y_{it}-Y_{Pit})}{Y_{it}} \qquad (3.17)$$

and the root mean square percent error (MSPE):

$$MSPE = \sqrt{1/T \sum_{t=1}^{T} [\frac{Y_{it}-Y_{iP}}{Y_{it}}]^2} \qquad (3.18)$$

The use of percentage changes is preferred to the first differences, since the former has the advantage of facilitating the comparison among different variables. On the other hand, where the variables experience growth or increasing trend, it puts the most recent changes more nearly on the same level as earlier changes. Although the use of logarithmic differences tend to make symmetrical decreases and increases, the changes are preferred because they are arithmetically simpler.

It is possible, using the percentages and absolute error, to construct a similar measure to the inequality coefficient proposed by Theil (3.13). The ratio of mean absolute error to the mean absolute actual percentage change has this similitude:

$$\frac{1/T\Sigma|Y_{it}-Y_{Pit}|}{1/T\Sigma|\frac{Y_{it}-Y_{it-1}}{Y_{it}}|} = \frac{\Sigma|Y_{it}-Y_{Pit}|}{\Sigma|\frac{Y_{it}-Y_{it-1}}{Y_{it}}|}$$

It measures the size of the error relative to the magnitudes that are being predicted.

5. Control chart:

The control chart is based on the assumption that the sum of the forecast errors should approach zero. The plot of the cumulative forecast errors can give an idea of how the model is an adequate representation of reality. Platt (31, p. 598) points out:

> The non-zero sum of forecast errors would indicate either the choice of an inappropriate model to represent the systematic variations in the variables or as the result of shifts in the way certain variables are related. . . .

The control chart is the plot of the accumulative sum of forecast errors against time. It is possible to draw a confidence band around zero; this confidence band is given by:

$$K\sqrt{1/T \sum_{t=1}^{T} (e_t-\bar{e})^2} \qquad (3.19)$$

where:

K: is a predetermined integer.

6. Spectral analysis:

This is an infrequently used technique for evaluating econometric models. Spectral analysis can be used to obtain a frequency decomposition of the variance or covariance of an univariate or bivariate stochastic process respectively. A time series is considered as the observed behavior of a stochastic process during some arbitrary time intervals. Thus, it is possible to use the procedure applied by Naylor et al. (28, p. 333). This procedure is based on comparing the estimated spectrum of a series generated by simulative experiment with the estimated spectrum of the actual series as a mean of verifying the results of simulation.

Another possibility developed by Howrey (17, p. 75) is to derive the implied spectrum directly from the model. This technique avoids making the computation needed to obtain the simulated series.

Many economic variables are autocorrelated and intercorrelated, i.e., correlated with other economic variables. Simulated values of economic variables are also autocorrelated and intercorrelated. Spectral analysis can be used for analyzing intercorrelated and autocorrelated data by comparing their spectra.

CHAPTER IV. DETERMINATION OF WEIGHTS

To weight the observations allows us to assign attention to the measures of the independent variables, so that the prediction is hopefully improved. Assigning different weights to different observations results from believing that some observations are more important than others.

Several authors point out that the placing of more weight in the observations correspondent to those periods in which economic changes are unusual would improve the ability of the econometric model for forecasting, i.e. Howrey et al. (18).

Thus, the assignment of weight to each observation is of capital importance. It is possible to see two cases in the weight assignment:

1. The case in which the weights assigned are a function of the residuals.
2. The case in which the weights assigned are independent of the residuals.

Case no. 1 can be seen as implicit in the estimation procedure, and/or according to different assumptions on the model. The five cases of weighted regression considered in Chapter II can be placed in this category.

Thus, a common aspect to the estimation of models with a noncommon variance (heteroscedasticity), errors in variables and random coefficients, is the fact that the value of the observations is weighted inversely proportional to the size of the error variance; hence the weights appear as a function of the errors. This is easily observed in the formulas (2.15, 2.23 and 2.26). the other two cases, Case B and D, are also assigned to this category because the procedure of estimation is accomplished by using an iterative process, once more, the weights assigned are a function of the errors. Even in the procedures developed by Huber (19) in which he suggests to minimize the expression (2.18); and in the definition of weights proposed by Tukey and used by Fair (2.21) the assignment of weights is not independent of the residuals.

The second case is based on the premise that it is not necessary that large residuals should be treated with different weight than small ones.

The condition of independence among errors and weights suggests a "different" kind of weighting of the observations. One simple possibility would be to assign weights at random, but obviously this seems to be also one of the most unpropitious. This kind of assigning is not convenient because:

1. Each selection of random numbers is a selection of weights.

2. It would be necessary to make a decision about the correspondence between random number (weight) and observation.

These two facts are contrary to the idea that specific observations are more important than others.

Based on the effects of a change in the output and input price on the profit maximizing level of output of a competitive, single-output, multiple-input firm, Ladd (25, p. 9) finds that coefficients to be estimated for use in making the forecast are functions of current prices. From this he suggests that:

> If we are in period n and want to make forecasts for n+1, the "current condition" of the period n are the most important conditions to use. Sample periods in which the conditions were close to conditions in period n ought to be more important than sample periods in which conditions were greatly different from conditions in period n. . . .

This consideration leads to considering two possible measures of proximity as weights:

1. Temporal distance, where the weights are assigned by:

$$W_i = \frac{1}{(t+1-i)^{1/2}}$$

where the denominator is the square root of the distance in the time of the observation from the period t + 1.

The most recent observation has more weight than the earlier observations. To the last observation of the sample period is always assigned a weight equal to one. The weights obtained by this form could be called temporal weights.

2. Metric distance, the weights are given by:

$$W_i = \frac{1}{d(i,t)}$$

for

$i<t$

where:

$$d(i,t) = [\Sigma_j (X_{ji}-X_{jt})^2]^{1/2}$$

being:

X_{it}: the i-th component of the vector of independent variables in the t-th sample period.

The weights assigned by using metric distance depend on the data.

For the last period the weight should satisfy:

$$W_t \geq \max[1/d(i,t)]$$

but, it would be an arbitrary value which could be larger than one. The weights obtained by using metric distance could be called metric weights.

In this way, the temporal and metric distances allow us to weight the observations such that the residuals and weights are independent.

CHAPTER V. EMPIRICAL WORK

The empirical work in this thesis is based on the analysis of six different econometric models. They were developed and published by different authors. The selection of these models for comparing unweighted regression procedures with the weighted regression procedure for forecasting is quite arbitrary and the only reason for choosing these models was the availability of the data.

The empirical work includes three aspects:

1. The analysis of selected model, which involves:
    a. A general explanation of the purposes for constructing the model.
    b. Specification of the model:
        (1) Listing the variables explicitly included.
        (2) Stating the functional form of the equation.
        (3) The probability distribution of the error is assumed normal with mean zero and matrix of covariances $\sigma^2 I$.
    c. Data used in the estimation sample period.
2. Estimation of the selected model according to two different procedures:
    a. Using unweighted regression procedures for estimating the parameter in the model (OLS).

b. Using weighted regression procedures with two kinds of distances as weights:

(1) Temporal distances.

(2) Metric distances.

The metric distance considered as weight which was assigned to the last observation and which should satisfy:

$$W_t = \max 1/d(i,t)$$

was determined by adding the two largest weights from previous observations. The abbreviation TWR will be used to denote regression procedure using the inverse of the temporal distances as weights, and MWR to denote the regression procedure using the inverse of the metric distances as weights.

1. The assumptions about the error distribution in the unweighted regression procedure (OLS) allow derivation of confidence region and hypothesis testing for the estimators, several measures on goodness-of-fit, and tests against autocorrelation of the residuals.

Let W be the weighting matrix, then a weighted estimate of β is obtained by minimizing $e'We$ in the model $Y = X\beta+\varepsilon$ with

$$E(\varepsilon\varepsilon') = \sigma^2 I$$

So,

$$e'We = (Y-Xb_w)'W(Y-Xb_w)$$

where:

$W = \text{diag}(W_1, W_2, \ldots, W_t)$

The normal equations are:

$$X'WX\beta_W = X'WY \tag{5.1}$$

This can be written as:

$$(W^{1/2}X)'(W^{1/2}X)\beta_W = (W^{1/2}X)'(W^{1/2}Y) \tag{5.2}$$

The estimator of β_W is:

$$b_W = (X'WX)^{-1}X'WY \tag{5.3}$$

b_W: is unbiased and consistent estimator of β with covariance matrix equal to:

$$\text{Var}(b_W) = \sigma^2(X'WX)^{-1}X'WW'X(X'WX)^{-1} \tag{5.4}$$

It is not easy to derive the sampling distribution of these estimators because they are a function of the assigned weight, hence the use of t and F ratios for construction of confidence intervals and tests of significant is not valid. Thus, the ratios of TWR and MWR coefficients to standard errors which are shown in several tables should be considered as descriptive statistics.

From Equation 5.2, it is clear that obtaining the estimator is accomplished by multiplying the i-th row of [Y X] by the weight assigned to the i-th observation and running the regression. The data transformation eliminates the intercept term creating a new variable, namely $X_{1i} = W_i$. Since the regression equation is forced through the origin, the computed residuals are uncorrelated with the explanatory variables; but it need not be true that $\Sigma e_i = 0$ because in that kind of model the sum of the residual ($\Sigma e_i = 0$) is not one of the normal equations and the partition of the total sum of the squares about the mean into the total sum of squares due to regression plus the residual sum of squares no longer holds, in general. From this, comparison of R^2 for the two regression procedures is invalidated.

2. The procedures for evaluating a set of forecasts developed in Chapter II are of little assistance for judging the results obtained in this thesis. The reason for this affirmation is based on the way in which the forecasts are constructed. Since each forecast is one-step-ahead forecast, this implies that the weight assigned to a specific observation is different for each sample period; hence, each observation added to the data changes the sample period. Therefore, the temporal and metric weights also change. This procedure violates the assumption that the forecast values are generated by the same structure.

This reason leads to designing new measures for comparing the forecasts constructed by using different regression procedures.

Let $e_{ijk} = A_{ijk}-F_{ijk}$ be a forecast error of i-th period (i = 1,2), using the j-th (j = 1,2,3) regression procedure in the k-th econometric model ($k = 1,2,3,4,5,6$). Thus $j = 1$ corresponds to OLS, $j = 2$ denotes TWR and $j = 3$ corresponds to MWR.

1. A first measure may be based on reducing forecast errors to a common value by dividing the forecast error by the variance of the endogenous variable and taking the arithmetic mean. This

$$\text{a.}\quad M_{jk} = \frac{1}{2}\left[\frac{e^2_{1jk}}{\sigma^2_{11k}} + \frac{e^2_{2jk}}{\sigma^2_{21k}}\right] \qquad (5.5)$$

for

$j = 1,2,3$

$k = 1,2,3,4,5,6$

where

$$\sigma^2_{11k} = \frac{1}{t-1}\sum_{i=1}^{t}(Y_i-\overline{Y})^2$$

and

$$\sigma^2_{21k} = \frac{1}{t}\sum_{i=1}^{t+1}(Y_i-\overline{Y})^2$$

M_{jk} will be equal to zero when $e_{1jk} = e_{2jk} = 0$ ($\sigma^2 < \infty$), that is to say, when the j-th procedure in the k-th econometric model gives a perfect forecast.

So, smaller values of M_{jk} are preferred over larger ones.

b. The geometric mean of the ratios e_{ijk}^2/σ_{i1k}^2 would seem to be an adequate measure for averaging these quantities. Thus, it is possible to define:

$$G_{jk} = \sqrt{\left(\frac{e_{1jk}^2}{\sigma_{11k}^2}\right)\left(\frac{e_{2jk}^2}{\sigma_{21k}^2}\right)} \qquad (5.6)$$

for $j = 1,2,3$

$k = 1,2,3,4,5,6$

The use of a geometric mean is justified because it is not so heavily weighted by extreme values as is the arithmetic mean.

The interpretation of G_{jk} is similar to the M_{jk}.

2. Several others measures can be designed for comparing the forecast obtained by using unweighted regression with those constructed from weighted regression.

a. The ratios of average square forecast error can be used for comparing the different forecasts with the forecast using unweighted regression.

$$Z_{ijk} = \frac{e_{1jk}^2 + e_{2jk}^2}{e_{11k}^2 + e_{21k}^2} \qquad (5.7)$$

for $j = 2,3$

$k = 1,2,3,4,5,6$

$Z_{12k} < Z_{11k}$ and $Z_{13k} < Z_{11k}$ indicate that the weighted regression procedures perform better than the unweighted one, also, the smaller value for Z_{1jk} would indicate which regression procedure in particular (TWR or MWR) is better in terms of forecasting.

b. A measure quite similar to formula (5.7) is the ratios of absolute average forecast errors. This measure has the advantage that large errors are not compensated disproportionately at the cost of smaller ones.

Thus,

$$A_{1jk} = \frac{|e_{1jk}| + |e_{2jk}|}{|e_{11k}| + |e_{21k}|} \qquad (5.8)$$

for

$j = 2,3$

$k = 1,2,3,4,5,6$

The interpretation of this measure is the same as (5.3).

3. An overall evaluation of the forecasts using different regression procedures can be accomplished by using the results obtained by applying the formulas 5.5, 5.6, 5.7 and 5.8 to each econometric model.

a. Defining TM_j as:

$$TM_j = \frac{1}{6} \sum_{k=1}^{6} M_{jk} \qquad (5.9)$$

for

$j = 1,2,3$

where M_{jk} is given by 5.5.

The decision rule applied to this measure is that the smaller value of TM_j $(j = 1,2,3)$ will indicate that the procedure j has a better overall performance for forecasting.

The same statement can be applied to the following measures:

$$\text{b.} \quad TG_j = \sqrt[6]{\prod_{k=1}^{6} G_{jk}} \qquad (5.10)$$

for

$j = 1,2,3$

$$\text{c.} \quad TZ_{ij} = \frac{1}{6} \sum_{k=1}^{6} Z_{ijk} \qquad (5.11)$$

for

$j = 1,2,3$

and

$$\text{d.} \quad TA_{1j} = \frac{1}{6} \sum_{k=1}^{6} A_{1jk} \qquad (5.12)$$

for

$j = 1,2,3$

where G_{jk}, Z_{1jk} and A_{1jk} are given by the formulas 5.6, 5.7 and 5.8 respectively.

4. Based on the material exposed in Chapter II, it is possible to apply a statistical test for evaluating the capacity of the model to predict turning points. A (2 x 2) contingency table can be constructed and a Chi-square test used for testing the null hypothesis for independence between

actual and predicted turning points for each regression procedure.

The turning points may be arranged in the following 2 x 2 contingency table.

Number of actual and predicted turning points method of estimation j-th (for j = 1,2,3):

	Predicted turning points	Predicted no turning point	Total
Actual turning points	a	b	n_1
Actual no turning points	c	d	n_2
	m_1	m_2	N

There are two forecasts for each model, so the total number of observations ($N = n_1+n_2$) is going to be 12.

The test statistic is given by:

$$T = \frac{N(ad-bc)}{n_1 \ n_2 \ m_1 \ m_2} \tag{5.13}$$

The decision rule is to reject the null hypothesis if T exceeds the $1-\alpha$ quantile of a Chi-square random variable with one degree of freedom.

A measure for expressing the degree of dependence shown in a particular contingency table can be the Phi-coefficient defined as:

$$\Pi = \frac{ad-bc}{\sqrt{n_1 \times n_2 \times m_1 \times m_2}} \tag{5.14}$$

$$-1 \leq \Pi \leq 1$$

which is a special case of the Pearson product moment correlation coefficient.

Model A

This model was developed by Ryan and Abel (33, p. 105) for estimating the acreage planted in oats and effects of the U.S. government commodity programs to limit output.

The model may be expressed as:

$$A = f(P_S, P_L, V)$$

where:

A: is the acreage planted in oats

P_S: is the support price weighted by planting restriction

P_L: is the payment for diverting land from oat production

V: includes other variables and random factors

The variables explicitly included are:

Y: U.S. acreage of oats planted (in thousands)

X_2: U.S. average oats loan rate weighted by acreage restriction requirements (dollar per bushel)

X_3: U.S. acreage of wheat planted (in thousands)

X_4: U.S. acreage diverted under wheat programs (in thousands)

X_5: 0 in 1956-67 and 1 in 1968-69, to account for a change in the economy policy

X_6: linear trend

X_7: X_6 squared

The functional form of the equation is:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \varepsilon$$

The sample period for the 1970 forecast is from 1956 through 1969. In order to forecast 1971 the data used was 1956-1970.

The weights assigned to this model were constructed as explained in Chapter IV, and they are shown in Table 1a. It is interesting to note that the weights assigned to the observation using the inverse of the metric weights are truly irregular in size; they do not follow a specific pattern, and the idea of assigning more weights to more recent observations does not appear to be in agreement with the empirical results. Also, the metric weights are really small when they are compared with the temporal weights.

1. Estimates from the regression:

The estimated coefficients using unweighted and weighted regression procedures for both sample periods are shown in Tables 1b and 1c, respectively.

There are several aspects which are quite interesting in the values of the coefficients:

Table 1a. Weights (Model A)

Time	Data: 56-69		Data: 56-70	
	Temporal	Metric[a]	Temporal	Metric[a]
1956	0.267260	0.078119	0.258200	0.051904
1957	0.277350	0.210500	0.267260	0.342272
1958	0.288680	0.165158	0.277350	0.081436
1959	0.301510	0.088011	0.288680	0.057871
1960	0.316230	0.089946	0.301510	0.060210
1961	0.333333	0.089353	0.316230	0.059218
1962	0.353550	0.199155	0.333333	0.199807
1963	0.377960	0.249600	0.353550	0.107041
1964	0.408250	0.162334	0.377960	0.081484
1965	0.447210	0.201134	0.408250	0.086308
1966	0.500000	0.356502	0.447210	0.112613
1967	0.577350	0.057172	0.500000	0.041462
1968	0.707110	0.072440	0.577350	0.049062
1969	1.000000	0.606100	0.707110	0.150561
1970	-	-	1.000000	0.542100

[a]Metric value has been multiplied by 10^3.

Table 1b. Estimated coefficients according to different regression procedures (Model A)

Variables	Data: 1956-69		
	Unweighted	Weighted	
		Temporal	Metric
X_1	53,001.672** (8,568.388)	53,203.515** (7,630.429)	53,013.869** (6,656.571)
X_2	13,567.821* (4,320.741)	13,481.959* (4,314.127)	12,510.671* (4,321.027)
X_3	-0.231 (0.142)	-0.234 (0.127)	-0.219 (0.115)
X_4	-0.123 (0.113)	-0.127 (0.103)	-0.107 (0.095)
X_5	-24,009.988** (803.244)	-24,010.833** (772.333)	-24,238.993** (677.529)
X_6	-3,562.153** (427.267)	-3,564.825** (399.006)	-3,615.047** (322.160)
X_7	122.541* (36.899)	122.859** (33.774)	126.932** (26.531)

*$p < 0.05$.

**$p < 0.01$.

Table 1c. Estimated coefficients according to different regression procedures (Model A)

Variables	Data: 1956-70		
	Unweighted	Weighted	
		Temporal	Metric
X_1	53,776.419**	54,407.040**	55,214.556**
	(7,869.915)	(7,173.535)	(7,101.071)
X_2	13,470.664*	13,340.131*	12,794.534*
	(4,076.813)	(4,139.377)	(4,288.326)
X_3	-0.243	-0.251	-0.255
	(0.131)	(0.120)	(0.121)
X_4	-0.125	-0.129	-0.124
	(0.106)	(0.100)	(0.101)
X_5	-23,954,424**	-23,976.255**	-24,261.414**
	(746.732)	(745.866)	(563.302)
X_6	-3,597.524**	-3,626.194**	-3,746.514**
	(394.304)	(374.012)	(340.103)
X_7	125.651**	128.178**	137.060**
	(34.023)	(31.690)	(29.199)

*$p < 0.05$.

**$p < 0.01$.

a. The three estimates of a specific coefficient do not differ greatly in value. There is not a big difference among the coefficients according to different regression procedures.

In the 1956-69 sample period, the largest percentage variation is in the coefficients of X_2 estimated by using OLS and MWR. This difference amounts to 7.8 percent. The smallest percentage difference is in coefficients of X_5 estimated by OLS and TWR. The difference is 0.003 percent.

For the sample period 1956-70, the largest percentage difference is in the X_7 coefficients which for OLS is 125.651 and for MWR is 137.060; this difference amounts to 9.1 percent. The smallest percentage is for X_5's coefficients with a variation of 0.9 percent between the OLS and TWR procedures.

The average percentage of change of the coefficients between the two sample periods is 1.3 percent for the coefficients estimated using OLS, 1.6 percent for those estimated by TWR and 3.6 percent for the coefficients obtained by using MWR. The smallest and largest changes are in X_3's and X_5's coefficients estimated by MWR, the difference amounts to 0.1 percent and 16.4 percent respectively.

The sign attached to each coefficient for different regression procedures is the same; that is, the direction of the change of the dependent variable in response to a unit change in a particular independent variable, holding constant the level of other independent variables, is the same whatever the regression procedure is.

b. When an unweighted regression procedure is used, the standard errors (s.e.) of the estimators tend to be smaller than those obtained when a weighted regression procedure is used.

2. Tables 1d and 1e contain several statistics from the regression output. Although the comparison of R^2 for the two regression procedures is invalidated, the fit of the equations can be appreciated comparing the actual values and their correspondent fitted series over the sample period for each regression procedure. All three regression procedures predict the same quantity of turning points (1) of a total of 3 turning points for both of the sample periods.

The size of the Durbin-Watson statistic for OLS indicates negative serially correlated disturbance at the 0.05 level of significance.

The transformation of the variables, as they are affected by weights, leads to a decrease of the estimator of the variance (σ^2). The value of $\hat{\sigma}^2$ in MWR is the smaller one.

The plot of residuals against the predicted values, for all different procedures, does not indicate abnormality. So, the Least Squares analysis does not appear to be invalidated.

Table 1d. R^2, $\bar{R}^2$, Durbin-Watson and standard error according to different regression procedures (Model A)

Statistics	Data: 1956-69		
	Unweighted	Weighted	
		Temporal	Metric
R^2	0.996**	-	-
$\bar{R}^2$	0.985	-	-
DW	3.637	3.588	3.719
RHO[a]	-0.828	-0.804	-0.866
SE	622.295	361.513	6.598

[a]RHO = first order autocorrelation coefficient.

**$p < 0.01$.

Table 1e. R^2, $\bar{R}^2$, Durbin-Watson and standard error according to different regression procedures (Model A)

Statistics	Data: 1956-70		
	Unweighted	Weighted	
		Temporal	Metric
R^2	0.996**	-	-
$\bar{R}^2$	0.986	-	-
DW	3.642	3.613	3.711
RHO[a]	-0.836	-0.827	-0.878
SE	588.177	336.996	5.024

[a]RHO = first order autocorrelation coefficient.

**$p < 0.01$.

3. Analysis of forecast:

The actual values and the one-step-ahead forecast are shown in Table 1f. All the predicted values for 1970 under-

Table 1f. Actual and predicted values (Model A)

Time	Actual Value	Predicted Value		
		Unweighted	Temporal	Metric
1970	24,492	24,157.0	24,124.9	24,117.8
1971	21,926	22,108.8	22,128.6	22,213.2

estimate the actual value. On the other hand, all regression procedures overestimate the actual value for 1971.

The smallest underestimation and largest overestimation come from obtaining the forecast by using MWR.

There is no turning point for 1970 and the model, using the sample period 1956-69, does not predict any turning point. All of the regression procedures predict exactly one turning point for 1971.

The measures for accuracy of the forecast developed in this chapter indicate that the forecasts generated by using unweighted regression perform better than those obtained by using weighted regression procedures. This can be noted in Table 1g, the values for M and G were found by applying the formulas (5.5) and (5.6). The values of Z and A which compare the unweighted regression with the weighted one, assert a better performance of the former procedure for this model in particular.

Table 1g. Measures of accuracy forecast (Model A)

Measures	Unweighted	Weighted	
		Temporal	Metric
M	0.134[a]	0.161[a]	0.205[a]
G	0.113[a]	0.137[a]	0.199[a]
Z	-	1.207	1.527
A	-	1.100	1.277

[a]Multiplied by 10^2.

Model B

This model is also developed by Ryan and Abel (33, p. 105) for estimating the acreage planted in barley. The variables explicitly included are:

Y: U.S. acreage of barley planted (in thousands)

X_2: U.S. average barley loan rate (plus direct support payments, 1963-65) weighted by acreage restriction requirements (dollars per bushel)

X_3: U.S. average oats loan rate weighted by acreage restriction requirements (dollars per bushel)

X_4: U.S. acreage of wheat planted (in thousands)

X_5: U.S. acreage diverted under wheat programs, in thousands

X_6: 0 in 1949-65 and 1 in 1966-70

The functional form can be written as:

$$Y = \sum_{j=1}^{6} \beta_j X_j + \varepsilon$$

The first sample period is from 1949-69, for forecasting 1970; in order to forecast 1971 the data used is 1949-70.

The weights assigned to this model are shown in Table 2a. The metric weights are really small and they are always smaller than the temporal weights and follow an irregular pattern.

1. Estimates from the regression:

In the first sample period, the largest variation is in the coefficients of X_5 estimated by using OLS and MWR (310%). For the second sample period the largest difference is also in the X_5 coefficients (264%) obtained by using OLS and MWR.

The smallest difference in the first period is in the X_1

Table 2a. Weights (Model B)

Time	Data: 49-69		Data: 49-70	
	Temporal	Metric[a]	Temporal	Metric[a]
1949	0.218220	0.031608	0.213200	0.026435
1950	0.223610	0.049238	0.218220	0.037224
1951	0.229420	0.037502	0.223610	0.030295
1952	0.235700	0.037348	0.229420	0.030198
1953	0.242540	0.036988	0.235700	0.029969
1954	0.250000	0.072275	0.242540	0.048981
1955	0.258200	0.084835	0.250000	0.055625
1956	0.267260	0.078119	0.258200	0.051904
1957	0.277350	0.210500	0.267260	0.342273
1958	0.288680	0.165158	0.277350	0.081436
1959	0.301510	0.088011	0.288680	0.057871
1960	0.316230	0.089947	0.301510	0.060210
1961	0.333333	0.089354	0.316230	0.059218
1962	0.353550	0.199165	0.333333	0.199817
1963	0.377960	0.249632	0.353550	0.107043
1964	0.408250	0.162349	0.377960	0.081486
1965	0.447210	0.201175	0.408250	0.086311
1966	0.500000	0.356837	0.447210	0.112624
1967	0.577350	0.057174	0.500000	0.041463
1968	0.707110	0.072440	0.577350	0.049062
1969	1.000000	0.606400	0.707110	0.150561
1970	-	-	1.000000	0.542100

[a]Metric value has been multiplied by 10^3.

coefficients. This variation amounts to 0.8 percent for the coefficients obtained by OLS and MWR. In the second sample period the coefficients of X_1 estimaed by OLS and TWR have the smallest difference (0.6%).

Comparing the coefficients for the same regression procedure between the two sample periods, a percentage of change is obtained for the OLS, between the first and the second sample periods, of 0.5 percent; for the TWR the percentage is 3.4 percent and for MWR the change amounts to 7.5 percent (see Tables 2b and 2c).

It is important to indicate a change in the sign associated to X_5 coefficient in the first sample period. The sign attached to this coefficient, when it is obtained by using TWR procedure, is positive, but using OLS and MWR the sign is negative. The situation in the sample period 1949-70 is different from this: all of the signs of the X_5's coefficients are negative.

The standard errors of the coefficients do not have a regular pattern; their value is not always smaller for a particular regression procedure than for others.

Tables 2d and 2e show some results from the regression output.

The OLS procedure exactly predicts over the first sample period two out of thirteen turning points. The TWR and MWR exactly predict one and zero turning points, respectively.

Table 2b. Estimated coefficients according to different regresion procedures (Model B)

Variables	Data: 1949-69		
	Unweighted	Weighted	
		Temporal	Metric
X_1	30,050.022** (2,255.273)	29,351.549** (2,574.054)	30,292.521** (2,928.123)
X_2	9,970.264** (2,251.301)	9,768.948** (2,359.938)	10,286.765** (2,745.559)
X_3	-18,273.009** (5,543.036)	-16,802.430* (6,182.066)	-12,615.873* (5,789.208)
X_4	-0.219** (0.045)	-0.226** (0.053)	-0.289** (0.060)
X_5	-0.020 (0.087)	0.010 (0.094)	-0.082 (0.080)
X_6	-2,015.287** (679.056)	-1,962.884** (572.404)	-1,017.512 (588.744)

*$p < 0.05$.

**$p < 0.01$.

Table 2c. Estimated coefficients according to different regression procedures (Model B)

Variables	Data: 1949-70 Unweighted	Weighted Temporal	Weighted Metric
X_1	30,107.822** (2,160.345)	29,917.427** (2,433.440)	30,369.709** (2,540.569)
X_2	9,725.062** (2,166.362)	9,489.262** (2,200.799)	9,259.782** (1,460.272)
X_3	-18,256.214** (5,371.241)	-16,726.371* (5,914.457)	-14,014.748* (5,246.503)
X_4	-0.219** (0.043)	-0.227** (0.050)	-0.258** (0.054)
X_5	-0.025 (0.078)	-0.023 (0.079)	-0.091 (0.062)
X_6	2,049.698** (627.707)	-2,066.757** (543.194)	-1,736.257** (549.236)

*$p < 0.05$.

**$p < 0.01$.

Table 2d. R^2, $\bar{R}^2$, Durbin-Watson and standard error according to different regression procedures (Model B)

Statistics	Data: 1949-69		
	Unweighted	Weighted	
		Temporal	Metric
R^2	0.876**	-	-
$\bar{R}^2$	0.690	-	-
DW	1.982	2.070	1.861
RHO[a]	-0.066	-0.101	-0.012
SE	1,071.824	654.603	11.191

[a]RHO = first order autocorrelation coefficient.

**$p < 0.01$.

Table 2e. R^2, $\bar{R}^2$, Durbin-Watson and standard error according to different regression procedures (Model B)

Statistics	Data: 1949-70		
	Unweighted	Weighted	
		Temporal	Metric
R^2	0.880**	-	-
$\bar{R}^2$	0.704	-	-
DW	2.034	2.197	2.121
RHO[a]	-0.078	-0.157	-0.120
SE	1,038.772	612.777	8.321

[a]RHO = first order autocorrelation coefficient.

**$p < 0.01$.

In the second period the OLS exactly predicts one out of thirteen and TWR and MWR zero and zero turning points respectively.

The use of the Least Squares analysis does not appear to be invalidated: the plot of the residual against predicted values over the sample periods does not indicate any abnormality.

2. Forecast analysis:

Table 2f indicates that all of the three regression procedures overestimated the actual values for 1970 and 1971. The largest forecast error for 1970 corresponds to the forecasts of the MWR procedure, the smallest one corresponds to the forecast of the OLS procedure.

OLS and MWR procedures give the largest and smallest forecast error for 1971.

The M and Z measures indicate that the MWR procedure performs better than either OLS or TWR. In contrast, the OLS procedure is considered better according to measures G and A. Thus, for this model, there is an indeterminacy which procedure performs better (see Table 2g).

There is not any turning point for either 1970 or 1971 and the different regression procedures do not forecast any turning points for these two years.

Table 2f. Actual and predicted values (Model B)

Time	Actual Value	Predicted Value		
		Unweighted	Temporal	Metric
1970	10,435	10,663.4	11,052.6	11,277.152
1971	11,182	13,771.8	13,599.3	13,200.026

Table 2g. Measures of accuracy forecast (Model B)

Measures	Unweighted	Weighted	
		Temporal	Metric
M	0.023	0.022	0.017
G	0.004	0.011	0.012
Z	-	0.921	0.707
A	-	1.077	1.015

Model C1

This model was developed by Hun Lee Tong (41, p. 82). The purpose of the model is to estimate the housing demand and to evaluate the income and price elasticities of desired stock demand for nonfarm housing.

The variables explicitly included are:

Y: per family gross rate of nonfarm residential construction in real terms

X_2: Boeckh index of residential construction cost in real terms

X_3: per family nonfarm current income in the real terms derived from Raymond Goldsmith's series

X_4: product of contract interest rates and contract lengths on a sample of straight urban mortgage loans

X_5: loan-to-value ratios on a sample of straight urban mortgage loans

X_6: beginning-of-year per family nonfarm housing stock in real terms

The functional form of the equation regression is linear.

$$Y = \sum_{j=1}^{6} \beta_j X_j + \varepsilon$$

The first sample period goes from 1920 through 1939 and the second one from 1920 through 1940. The sample periods are used to forecast the one-step-ahead-forecast for 1940 and 1941 respectively.

The metric weights assigned to the observations of this model are small and follow an irregular pattern; the larger metric weights are not assigned to the most recent observations (see Table 3a). The metric weights are always smaller than the temporal weights.

1. Estimates from the regression:

The coefficients are shown in Tables 3b and 3c, for the first and second sample periods, respectively.

The largest and smallest variation for the same coefficient estimated using different procedures corresponds to the

Table 3a. Weights (Model C1)

Time	Data: 20-39		Data: 20-40	
	Temporal	Metric	Temporal	Metric
1920	0.223610	0.003378	0.218220	0.002681
1921	0.229420	0.003119	0.223610	0.002371
1922	0.235700	0.004061	0.229420	0.002895
1923	0.242540	0.013677	0.235700	0.007031
1924	0.250000	0.009903	0.242540	0.006601
1925	0.258200	0.005846	0.250000	0.004559
1926	0.267260	0.003600	0.258200	0.003493
1927	0.277350	0.002734	0.267260	0.002722
1928	0.288680	0.002313	0.277350	0.002251
1929	0.301510	0.001799	0.288680	0.001885
1930	0.316230	0.001779	0.301510	0.001742
1931	0.333333	0.001971	0.316230	0.001839
1032	0.353550	0.001691	0.333333	0.001493
1933	0.377960	0.001714	0.353550	0.001493
1934	0.408250	0.002143	0.377960	0.001815
1935	0.447210	0.003212	0.408250	0.002556
1936	0.500000	0.006141	0.447210	0.005094
1937	0.577350	0.009081	0.500000	0.007448
1938	0.707110	0.007602	0.577350	0.004313
1939	1.000000	0.022758	0.707110	0.009797
1940	-	-	1.000000	0.017246

Table 3b. Estimated coefficients according to different regression procedures (Model C1)

Variables	Data: 1920-39		
	Unweighted	Weighted	
		Temporal	Metric
X_1	817.604** (206.836)	808.900** (223.809)	713.346* (274.297)
X_2	-7.730** (2.525)	-6.592* (2.599)	-6.414* (2.946)
X_3	0.220** (0.062)	0.206** (0.062)	0.240* (0.083)
X_4	-2.764* (0.968)	-3.171** (0.954)	-2.496* (1.088)
X_5	8.102 (4.240)	8.176 (4.447)	6.446 (5.130)
X_6	-0.221** (0.046)	-0.238** (0.048)	-0.215** (0.056)

*$p < 0.05$.

**$p < 0.01$.

Table 3c. Estimated coefficients according to different regression procedures (Model C1)

Variables	Data: 1920-40		
	Unweighted	Weighted	
		Temporal	Metric
X_1	823.530** (199.890)	827.196** (218.276)	767.765** (260.191)
X_2	-7.784** (2.444)	-6.784* (2.566)	-6.669* (2.900)
X_3	0.221** (0.060)	0.211** (0.061)	0.243** (0.080)

*$p < 0.05$.

**$p < 0.01$.

Table 3c (Continued)

Variables	Data: 1920-40		
	Unweighted	Weighted	
		Temporal	Metric
X_4	-2.676** (0.904)	-2.802** (0.865)	-2.344* (0.948)
X_5	7.756 (3.989)	6.998 (4.136)	5.566 (4.632)
X_6	-0.218** (0.044)	-0.228** (0.047)	-0.214** (0.053)

coefficients associated with X_5 and X_1, respectively. This occurs in both sample periods. For the period 1920-39, the coefficient of X_5 estimated by OLS is 8.102 and when estimated by MWR is 6.446; this difference amounts to 20.4 percent. For the second period the variation of the same coefficients is 28 percent.

The smallest difference amounts to 1 percent and 0.4 percent for the first and second period respectively. These are for the TWR and OLS coefficients of X_1.

The average percentages of change of the coefficients between different sample periods are 1.18 percent for OLS, 3.2 percent for TWR and 3.3 percent for MWR.

The sign attached to each particular coefficient remains the same among regression procedures and sample periods.

The standard errors (s.e.) tend to be smaller for OLS than for TWR and MWR in both sample periods.

The OLS and TWR predict three out of four turning points

Table 3d. R^2, $\bar{R}^2$, Durbin-Watson and standard error according to different regression procedures (Model C1)

Statistics	Data: 1920-39		
	Unweighted	Weighted	
		Temporal	Metric
R^2	0.870**	-	-
$\bar{R}^2$	0.670	-	-
DW	1.441	1.549	1.370
RHO[a]	0.259	0.201	0.293
SE	33.447	19.041	2.476

[a]RHO = first order autocorrelation coefficient.

**$p < 0.01$.

and MWR procedure only predicts two of them over the first period. On the other hand, in the second sample period, all of the three procedures exactly predict three out of four turning points.

The value $\bar{R}^2$ is 23 percent less than the value of R^2, for both sample periods (see Tables 3e and 3f).

The plot of residuals against predicted value does not indicate abnormality.

Table 3e. R^2, $\bar{R}^2$, Durbin-Watson and standard error according to different regression procedures (Model Cl)

Statistics	Data: 1920-40 Unweighted	Weighted Temporal	Weighted Metric
R^2	0.870**	-	-
$\bar{R}^2$	0.676	-	-
DW	1.447	1.448	1.378
RHO[a]	0.261	0.251	0.294
SE	32.443	18.209	2.127

[a]RHO = first order autocorrelation coefficient.

**$p < 0.01$.

Table 3f. Actual and predicted values (Model Cl)

Time	Actual Value	Predicted Value Unweighted	Predicted Value Temporal	Predicted Value Metric
1940	140.1	154.611	165.755	162.163
1941	146.6	216.183	214.264	216.537

2. Forecast analysis:

No turning points exist beyond the sample period and no turning point is forecasted by any of the three regression procedures.

Table 3f shows that all three regression procedures overestimate the actual values for 1940 and 1941.

The largest forecast error for the actual value of 1940 corresponds to the predicted value using TWR procedure, the smallest one corresponds to OLS procedure. For the forecast of 1941, the MWR gives the largest forecast error and TWR the smallest one.

The measures of forecast accuracy indicate that the unweighted regression procedure performs better than the weighted regression procedure. The values obtained for M (formula 5.5) and G (formula 5.6) are smaller using OLS than using either TWR or MWR. Table 3g also shows that Z and A are greater than one. That indicates that OLS regression procedure gives better forecasts.

Table 3g. Measures of accuracy forecast (Model Cl)

Measures	Unweighted	Weighted	
		Temporal	Metric
M	0.414	0.427	0.440
G	0.162	0.278	0.247
Z	-	1.036	1.064
A	-	1.110	1.094

Model C2

This model was also developed by Hun Lee Tong, however, he made a different assumption, namely, that the housing demand is more responsive to permanent income than to current income. Thus, the variable X_3 is now per-family permanent income, derived from Friedman's per capita permanent income series.

The sample periods are still the same as for model C1.

The weights assigned to this model are shown in Table 4a. The average of the metric weights assigned to C1 is smaller than the average of the metric weights for the model C2 in both sample periods.

1. Estimates from the regression:

The coefficients according to different regression procedures for the first and second sample periods are shown in Tables 4b and 4c, respectively. For the first sample period, the largest difference is between the coefficients of X_1 estimated by MWR and OLS (26%). The smallest difference for the same sample period corresponds to the X_4 coefficient: -2.998 for OLS and -2.995 for TWR, the percentage of change is 0.1 percent.

In the second sample period, the largest difference is in the coefficients of X_5 estimated by MWR and by OLS

Table 4a. Weights (Model C2)

Time	Data: 20-39		Data: 20-40	
	Temporal	Metric	Temporal	Metric
1920	0.223610	0.004024	0.218220	0.003539
1921	0.229420	0.007351	0.223610	0.005052
1922	0.235700	0.006434	0.229420	0.004621
1923	0.242540	0.010022	0.235700	0.006420
1924	0.250000	0.009872	0.242540	0.007468
1925	0.258200	0.005842	0.250000	0.005273
1926	0.267260	0.003612	0.258200	0.003520
1927	0.277350	0.002701	0.267260	0.002700
1928	0.288680	0.002244	0.277350	0.002241
1929	0.301510	0.001847	0.288680	0.001880
1930	0.316230	0.001685	0.301510	0.001699
1931	0.333333	0.001924	0.316230	0.001918
1932	0.353550	0.002190	0.333333	0.002095
1933	0.377960	0.002399	0.353550	0.002198
1934	0.408250	0.002649	0.377960	0.002364
1935	0.447210	0.003359	0.408250	0.002868
1936	0.500000	0.005448	0.447210	0.004318
1937	0.577350	0.009393	0.500000	0.006793
1938	0.707110	0.015932	0.577350	0.008263
1939	1.000000	0.025954	0.707110	0.015226
1940	-	-	1.000000	0.023490

Table 4b. Estimated coefficients according to different regression procedures (Model C2)

Variables	Data: 1920-39		
	Unweighted	Weighted	
		Temporal	Metric
X_1	860.908** (227.878)	897.823** (224.241)	636.218* (240.865)
X_2	-8.916** (2.671)	-8.400** (2.450)	-7.824** (2.555)
X_3	0.332* (0.118)	0.317* (0.108)	0.432** (0.132)
X_4	-2.998* (1.061)	-2.995* (1.050)	-2.696* (0.997)
X_5	10.595* (4.381)	9.833* (4.431)	8.613 (4.347)
X_6	-0.321** (0.063)	-0.326** (0.059)	-0.315** (0.063)

$^{*}p < 0.05$.

$^{**}p < 0.01$.

Table 4c. Coefficients according to different regression procedures (Model C2)

Variables	Data: 1920-40		
	Unweighted	Weighted	
		Temporal	Metric
X_1	893.892** (222.087)	936.825** (225.951)	749.813** (253.320)
X_2	-9.267** (2.611)	-8.745** (2.524)	-8.200** (2.757)
X_3	0.320* (0.116)	0.308* (0.110)	0.401* (0.137)
X_4	-2.838* (1.033)	-2.609* (1.014)	-2.058 (1.024)
X_5	10.117* (4.299)	8.723 (4.370)	6.698 (4.498)
X_6	-0.308** (0.060)	-0.310** (0.059)	-0.299** (0.065)

*p < 0.05.

**p < 0.01.

Table 4d. R^2, $\bar{R}2$, Durbin-Watson and standard error according to different regression procedures (Model C2)

Statistics	Data: 1920-39		
	Unweighted	Weighted	
		Temporal	Metric
R^2	0.844**	-	-
$\bar{R}^2$	0.610	-	-
DW	1.353	1.299	1.296
RHO[a]	0.320	0.344	0.346
SE	36.792	19.961	2.481

[a]RHO = first order autocorrelation coefficient.

**p < 0.01.

Table 4e. R^2, $\bar{R}^2$, Durbin-Watson and standard error according to different regression procedures (Model C2)

Statistics	Data: 1920-40		
	Unweighted	Weighted	
		Temporal	Metric
R^2	0.837**	-	-
$\bar{R}^2$	0.601	-	-
DW	1.317	1.227	1.154
RHO[a]	0.328	0.368	0.400
SE	36.413	19.759	2.408

[a]RHO = first order autocorrelation coefficient.

**$p < 0.01$.

(33.8%). The smallest difference corresponds to the X_6 coefficients obtained by using OLS and TWR (0.6%).

The average percentages of change of the coefficients between sample periods are 1.2 percent for OLS, 4.7 percent for TWR and 4.9 percent for MWR.

The sign attached to a particular coefficient remains constant whatever the regression procedure and sample period are.

For the first sample period, 1920-39, the OLS and MWR procedures exactly predict three out of four turning points over the entire period, the TWR procedure exactly predicts all the turning points, but the TWR procedure for the same sample period also predicts six more turning points, which do not correspond with any in the actual series.

In the sample period 1920-40, the OLS exactly predicts all the four turning points in the actual series but it also predicts six more for a total of ten. The TWR and MWR predict two and three out of four turning points, respectively.

No abnormality is detected by the plotting of the residuals against predicted value for both of the sample periods.

2. Forecast analysis:

No turning points exist beyond the sample period and none are forecasted.

All of the three procedures overestimate the actual values for 1940 and 1941. The smallest forecast errors correspond to the forecast obtained by using TWR procedures for 1940 and 1941. The smallest forecast errors correspond to the forecast obtained by using TWR proceudres for forecasting both values. The largest forecast errors correspond to the forecast obtained by applying MWR and OLS for the 1940 and 1941 forecasts respectively (see Table 4f).

Table 4g, based on the formulas 5.5 to 5.8, indicates that the TWR procedure performs better than the other two regression procedures for forecasting the actual values.

Table 4f. Actual and predicted values (Model C2)

Time	Actual Value	Predicted Value		
		Unweighted	Temporal	Metric
1940	140.1	179.788	175.830	181.984
1941	146.6	213.397	201.352	210.722

Table 4g. Measures of accuracy forecast (Model C2)

Measures	Unweighted	Weighted	
		Temporal	Metric
M	0.490	0.346	0.475
G	0.425	0.314	0.431
Z	-	0.708	0.972
A	-	0.850	0.995

Model D

The objectives of this model were to provide a basis for appraising feed consumption under alternative programs and for projecting the demand for feed concentrates. The model was developed by Ahalt and Egbert (1, p. 41).

The model may be written as:

$$F_t = f(L_t, P_{Lt-i}, P_{ft-i}, U_t)$$

where:

F_t: feed used in year t

L_t: livestock production or inventory in year t

P_{Lt-i}: livestock prices in year t-i

P_{ft-i}: feed prices in year t-i

U_t: random factors

The variables explicitly included are:

Y: total concentrates feed (million tons)

X_2: total livestock production units (millions)

X_3: ratio of livestock and product prices to feed grains and hay prices, multiplied by 100.

The functional form of the equation is:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

The sample period for the 1961 forecast is 1947-1960. The forecast for 1962 is based on the sample period 1947-61.

Table 5a shows the weights assigned to this model.

The weights based on the metric distance have an increasing trend through the sample period. That implies that more weight is assigned to the last observations. The metric weights are smaller than the temporal weights. For the first sample period, the largest metric weight which is assigned to the last observation can only be compared with the temporal weights assigned to the two first observations. For the second period the metric weights increase in

Table 5a. Weights (Model D)

Time	Data: 47-60		Data: 47-61	
	Temporal	Metric	Temporal	Metric
1947	0.267260	0.014819	0.258200	0.014910
1948	0.277350	0.017428	0.267260	0.017531
1949	0.288680	0.032500	0.277350	0.031491
1950	0.301510	0.031969	0.288680	0.031956
1951	0.316230	0.034762	0.301510	0.034381
1952	0.333333	0.022029	0.316230	0.022217
1953	0.353550	0.021186	0.333333	0.021339
1954	0.377960	0.020423	0.353550	0.020839
1955	0.408250	0.023408	0.377960	0.024021
1956	0.447210	0.020934	0.408250	0.021377
1957	0.500000	0.037704	0.447210	0.038008
1958	0.577350	0.099381	0.500000	0.078266
1959	0.707110	0.176777	0.577350	0.158114
1960	1.000000	0.276158	0.707110	0.353553
1961	-	-	1.000000	0.511667

value quickly.

1. Estimates from the regression:

Tables 5b and 5c present the values of the estimated coefficients for different regression procedures for first and second sample periods, respectively.

For the sample period 1947-60 the largest difference

Table 5b. Estimated coefficients according to different regression procedures (Model D)

Variables	Data: 1947-60		
	Unweighted	Weighted	
		Temporal	Metric
X_1	-52.176**	-66.218**	-91.300**
	(16.100)	(17.899)	(18.378)
X_2	0.778**	0.863**	1.090**
	(0.123)	(0.141)	(0.153)
X_3	0.237**	0.228*	0.113
	(0.071)	(0.081)	(0.097)

*$p < 0.05$.

**$p < 0.01$.

Table 5c. Estimated coefficients according to different regression procedures (Model D)

Variables	Data: 1947-61		
	Unweighted	Weighted	
		Temporal	Metric
X_1	-64.109**	-79.290**	-104.152**
	(15.731)	(16.545)	(14.709)
X_2	0.853**	0.950**	1.175**
	(0.125)	(0.136)	(0.127)
X_3	0.227*	0.209*	0.094
	(0.077)	(0.085)	(0.092)

*$p < 0.05$.

**$p < 0.01$.

Table 5d. R^2, $\bar{R}^2$, Durbin-Watson and standard error according to different regression procedures (Model D)

Statistics	Data: 1947-60		
	Unweighted	Weighted	
		Temporal	Metric
R^2	0.950**	-	-
$\bar{R}^2$	0.885	-	-
DW	1.840	1.804	1.639
RHO[a]	-0.058	-0.043	0.080
SE	3.188	2.272	0.780

[a]RHO = first order autocorrelation coefficient.

**$p < 0.01$.

Table 5e. R^2, $\bar{R}^2$, Durbin-Watson and standard error according to different regression procedures (Model D)

Statistics	Data: 1947-61		
	Unweighted	Weighted	
		Temporal	Metric
R^2	0.953**	-	-
$\bar{R}^2$	0.893	-	-
DW	1.557	1.516	1.381
RHO[a]	0.127	0.142	0.242
SE	3.445	2.314	0.772

[a]RHO = first order autocorrelation coefficient.

**$p < 0.01$.

among the coefficients for the different regression procedures is in X_1's coefficients, the variation between the OLS and MWR coefficients amount to 75 percent. The smallest variation is in X_3's coefficients related by OLS and TWR (3.4%). This situation is also present for the sample period 1947-61; the variation amounts to 62.5% and 7.9 percent, respectively.

The average percentage of change in the coefficients between the two sample periods is 9.7 percent for the unweighted regression procedure, 11.8 percent for the TWR and 24.3 percent for MWR.

The sign attached to each coefficient is constant whatever the regression procedure and the sample period.

The standard errors (s.e.) tend to be smaller for OLS than either TWR and MWR in both sample periods.

For the first sample period all three regression procedures predict four turning points out of six real turning points. In the second sample period only the MWR predicts three out of six actual turning points, OLS and TWR predict four turning points.

The plot of residuals against the predicted values, for different procedures, does not indicate abnormality; so, the Least Squares Analysis does not appear to be invalidated.

2. Forecast analysis:

There is no turning point in 1961, none of the regression procedures forecast any turning points. Only MWR procedure forecasts the turning point for 1962.

Table 5f presents the actual and forecast values for model D. The actual value for 1961 is underestimated for all the three regression procedures. In contrast, the actual value for 1962 is overestimated. The MWR gives the smallest forecast error for the predicted value of 1961, and OLS gives the largest one.

A contrary situation is given when the value for 1962 is forecast. The MWR procedure gives the largest forecast error and OLS gives the smallest one.

The results of applying the formulas for measuring the accuracy of the forecast are given in Table 5g. It is necessary to mention the following:

1. M indicates that the TWR performs better than the other two; on the other hand, G points to OLS as the best one.

2. The value of Z indicates a better performance of TWR than the OLS; and from the A value, it is possible to consider the OLS as superior to the other two regression procedures. Thus, under these measures it is not possible to say which regression procedure performs better in forecasting.

Table 5f. Actual and predicted values (Model D)

Time	Actual Value	Predicted Value		
		Unweighted	Temporal	Metric
1961	152.9	146.458	148.253	150.561
1962	152.0	152.994	155.241	158.497

Table 5g. Measures of accuracy forecast (Model D)

Measures	Unweighted	Weighted	
		Temporal	Metric
M	0.124	0.087	0.114
G	0.033	0.078	0.079
Z	-	0.755	1.122
A	-	1.061	1.188

Model E

The purpose of this model is to explain post-war state and local government new debt patterns. The author of this model is M. Tanzer (38, p. 237).

The variables included in the model are:

Y: state new debt, excluding the toll road sector (in billions of dollars)

X_2: state (nontoll) capital expenditures (less federal highway grants)

X_3: lagged stock of (nontoll) liquid assets

X_4: index of interest rate changes

The model can be written as:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

The sample periods are 1953-II to 1959-IV and 1953-II to 1960-I for the first and second forecasts respectively. The data are quarterly data.

The metric weights assigned to this model (see Table 6a) show an irregular pattern. The most recent observations are not affected by heavier weights but the weights tend to increase over time.

The weights constructed using the temporal distance are in most of the cases smaller than the weights based on metric distances.

1. Estimates from regression:

There is quite a big difference among the coefficients estimated (see Tables 6b and 6c).

For the sample period, 1953-II to 1960-I, the coefficient of X_1 of OLS has attached a contrary sign to the sign assigned to TWR and MWR.

In both sample periods, the largest difference between two estimates of a coefficient corresponds to the coefficient of X_1. The variation amounts to 62 percent and 290 percent for the coefficients estimated by OLS and TWR for the first and second periods respectively. The smallest difference,

Table 6a. Weights (Model E)

Time	Data: 53-II to 59-IV		Data: 53-II to 60-I	
	Temporal	Metric	Temporal	Metric
1953-II	0.192450	0.304671	0.188980	0.277052
-III	0.196120	0.333537	0.192450	0.328549
-IV	0.200000	0.325307	0.196120	0.359382
1954-I	0.204120	0.280492	0.200000	0.333196
-II	0.208510	0.241140	0.204120	0.294611
-III	0.213200	0.246805	0.208510	0.304488
-IV	0.218220	0.251044	0.213200	0.312192
1955-I	0.223610	0.311400	0.218220	0.385400
-II	0.229420	0.390882	0.223610	0.447841
-III	0.235700	0.451754	0.229420	0.431733
-IV	0.242540	0.399936	0.235700	0.338701
1956-I	0.250000	0.397260	0.242540	0.336976
-II	0.258200	0.407373	0.250000	0.342337
-III	0.267260	0.431046	0.258200	0.354706
-IV	0.277350	0.421911	0.267260	0.348924
1957-I	0.288680	0.405284	0.277350	0.338991
-II	0.301510	0.390423	0.288680	0.329701
-III	0.316230	0.428857	0.301510	0.350840
-IV	0.333333	0.531216	0.316230	0.482484
1958-I	0.353550	0.506078	0.333333	0.609201
-II	0.377960	0.392701	0.353550	0.549608
-III	0.408250	0.416114	0.377960	0.615073
-IV	0.447210	0.693092	0.408250	1.000150
1959-I	0.500000	1.183700	0.447210	0.771127
-II	0.577350	0.851102	0.500000	0.476163
-III	0.707110	0.942851	0.577350	0.490120
-IV	1.000000	2.126551	0.707110	0.987730
1960-I	-	-	1.000000	1.987880

Table 6b. Estimated coefficients according to different regression procedures (Model E)

Variables	Data: 1953-II to 1959-IV		
	Unweighted	Weighted	
		Temporal	Metric
X_1	0.194 (0.247)	0.074 (0.222)	0.111 (0.218)
X_2	0.432** (0.080)	0.476** (0.069)	0.465** (0.069)
X_3	-0.320** (0.074)	(0.297** (0.053)	-0.298** (0.051)
X_4	-0.115** (0.020)	-0.122* (0.020)	-0.122** (0.021)

** $\underline{p} < 0.01$.

Table 6c. Estimated coefficients according to different regression procedures (Model E)

Variables	Data: 1953-II to 1960-I		
	Unweighted	Weighted	
		Temporal	Metric
X_1	0.073 (0.237)	-0.139 (0.224)	-0.116 (0.217)
X_2	0.465** (0.077)	0.532** (0.072)	0.525** (0.070)
X_3	-0.268** (0.066)	-0.221** (0.050)	-0.206** (0.049)
X_4	-0.112** (0.021)	-0.111** (0.021)	-0.110** (0.222)

** $\underline{p} < 0.01$.

between OLS and a weighted procedure, corresponds to X_4's coefficients when they are estimated by TWR (6% and 1%).

The largest variation for the same coefficient in a different sample period is in the X_1's coefficient obtained by MWR, which changes from 0.111 in the first period to -0.116 in the second period (-2.290%). The smallest variation corresponds to the X_4's coefficient when using OLS procedure. The change is from -0.115 to -0.112 in the second period; this represents a variation of 3 percent.

The average percentage of change in the coefficients between the sample periods is 12 percent for OLS, 30 percent for TWR and 53 percent when MWR procedure is used.

The standard errors of the coefficients tend to be smaller for OLS.

Tables 6d and 6e show some results from the regression output.

In the first sample period the OLS predicts exactly three out of ten turning points, the weighted regression procedures only predict exactly two of them over the sample period 1953-II to 1959-IV. For the second sample period all of the three regression procedures exactly predict only two out of eleven turning points.

The plot of residuals against predicted values, in both sample periods and regression procedures does not indi-

Table 6d. R^2, $\bar{R}^2$, Durbin-Watson and standard error according to different regression procedures (Model E)

Statistics	Data: 1953-II to 1959-IV		
	Unweighted	Weighted	
		Temporal	Metric
R^2	0.920**	-	-
$\bar{R}^2$	0.826	-	-
DW	0.918	1.074	1.015
RHO[a]	0.511	0.429	0.461
SE	0.146	0.075	0.091

[a]RHO = first order autocorrelation coefficient.

**$\underline{p} < 0.01$.

Table 6e. R^2, $\bar{R}^2$, Durbin-Watson and standard error according to different regression procedures (Model E)

Statistics	Data: 1953-II to 1960-I		
	Unweighted	Weighted	
		Temporal	Metric
R^2	0.919**	-	-
$\bar{R}^2$	0.825	-	-
DW	0.903	1.026	1.094
RHO[a]	0.513	0.438	0.411
SE	0.149	0.080	0.098

[a]RHO = first order autocorrelation coefficient.

**$\underline{p} < 0.001$.

cate abnormality; so, the Least Squares Analysis does not appear to be invalidated.

2. Forecast analysis:

No regression procedure forecasts the turning point for 1960-I. All three regression procedures overestimate the actual values for 1960-I and 1960-II. The OLS gives the largest forecast error (in absolute value), and MWR gives the smallest one for both sample periods (see Table 6f).

The measures of accuracy developed in this chapter indicate that the MWR performs better than either OLS or TWR (see Table 6g).

Table 6f. Actual and predicted values (Model E)

Time	Actual Value	Predicted Value		
		Unweighted	Temporal	Metric
1960-I	1.86	2.108	2.100	2.098
1960-II	1.79	2.082	2.024	1.998

Table 6g. Measures of accuracy forecast (Model E)

Measures	Unweighted	Weighted	
		Temporal	Metric
M	0.304	0.233	0.208
G	0.300	0.232	0.205
Z	-	0.765	0.681
A	-	0.878	0.826

7. Overall appraisal:

After analyzing each model according to the different regression procedures, it is necessary to evaluate the overall performance of each regression procedure.

1. Weights:

The weights assigned to the observation have two characteristics:

a. Temporal weights increase over time and tend to be larger than metric weights.

b. Metric weights follow an irregular pattern; that means that the more recent observations are not always assigned more weight.

2. Estimates from the regression:

On this aspect, it is possible to point out:

a. There is not a big difference among the estimated coefficients according to different regression procedures, for the same sample period and the same model.

b. The largest differences between the coefficients estimated by using OLS and a weighted regression procedure correspond to coefficients estimated by using MWR (nine out of twelve). Seven out of nine largest variations are in coefficients of X_1.

The smallest variations correspond to the coefficients estimated by applying OLS and those estimated by using TWR procedure (eleven out of twelve). There is not a particular coefficient associated to those changes.

c. The difference between coefficients for the two sample periods is 2.3 percent for the coefficient estimated by using OLS, 5.5 percent for those estimated by TWR, and 9.1 percent for the coefficients estimated by applying MWR procedure (see Table 7).

d. The standard errors of the estimated coefficients tend to be smaller when OLS is used than when a weighted regression procedure is utilized; this situation is present in four out of six models. In the models C2 and E the s.e. do not have a specific pattern.

e. In general, the direction of the change of the dependent variable, for a specific model, in response to a unit change in a particular independent variable, holding constant the level of other independent variables, is the same whatever the regression procedure is. The only

Table 7. Percentage of change between coefficients of two sample periods

Model	Procedure		
	OLS	TWR	MWR
A	1.3	1.6	3.6
B	0.5	3.4	7.5
C1	1.8	3.2	3.3
C2	1.2	4.7	4.9
D	9.7	11.8	24.3
E	12.0	30.0	53.0

exceptions to this statement are the models B and E, in which a particular coefficient has different signs for different regression procedures.

f. The data transformation did not significantly affect the correlation between the successive residuals. Fourteen out of twenty-four (58.3%) first order autocorrelation coefficient (RHO) values showed an increment, and only in one case (Model D Second Sample Period) RHO changed sign (positive to negative).

3. Of turning points over the sample period for the first sample periods, the OLS procedure exactly predicts 54 percent, TWR and MWR procedures exactly predict 52 percent and 30 percent respectively.

For the second sample periods, the OLS also exactly predicts the greater number of turning points: 51 percent, whereas TWR and MWR only predict 44 percent.

Table 8. Measures of overall appraisal of forecasts

Measures	Unweighted	Weighted	
		Temporal	Metric
TM	0.226	0.186	0.243
TG	0.038	0.054	0.128
TZ	-	0.899	1.012
TA	-	1.013	1.066

4. Forecast analysis:

The results of applying formulas 5.12, 5.13, .14 and 5.15 are shown in Table 8.

The values obtained for TG and TA indicate that the unweighted regression procedure, OLS, performs better than either TWR or MWR procedures.

On the other hand, TM and TZ identify TWR as the procedure with better performance in forecasting.

Regression procedures overestimate ten out of twelve (83.3%) actual values. Four out of six largest errors correspond to forecasts obtained by using OLS, three of the smallest forecast errors correspond to predictions associated to MWR procedure (see Table 9).

The MWR forecasts two out of three turning points beyond the sample periods: OLS and TWR predict one each.

The test statistic (formula 5.13) for testing the null hypothesis for independence between actual and forecast

Table 9. Forecast errors for each model according different regression procedures

Model	Forecast	Procedure		
		OLS	TWR	MWR
A				
	1	335.00	367.10	374.16
	2	-182.80	-202.60	-287.15
B				
	1	-228.40	-617.60	-842.15
	2	-2,589.80	-2,417.30	-2,018.03
C1				
	1	-14.51	-25.65	-22.06
	2	-69.58	-67.66	-69.94
C2				
	1	-39.69	-35.73	-41.88
	2	-66.80	-54.75	-64.12
D				
	1	6.44	4.65	2.34
	2	-0.99	-3.24	-6.50
E				
	1	-0.25	-0.24	-0.24
	2	-0.29	-0.23	-0.21

turning points, beyond the sample period, for each regression procedure indicates that there is no reason for rejecting the null hypothesis at 5 percent level.

The measure for expressing the degree of dependence between predicted turning points and actual turning points is 0.52 for OLS, and 0.77 for TWR and MWR.

CHAPTER VI. CONCLUSIONS

One of the specific objectives of this thesis is to compare the unweighted and weighted regression and to evaluate their performance for forecasting. Before, making a decision about which regression procedure is better for forecasting, it is convenient to analyze several points:

1. The number of forecasts utilized is small. Only two forecasts were obtained from each econometric model, hence the number of turning points which could be predicted beyond the sample period is also small.

2. Measures of accuracy for forecasts do not clearly indicate which regression procedure is better.

It is convenient to recall that G and A are a geometric mean and ratios of absolute values respectively, both of these measures do not tend to be affected by large errors, (large under and/or overestimations). On the contrary, M and Z are quite affected by extreme values of the forecast error.

These two points help to understand situations such as model E and model B, in which it is not possible to determine which regression procedure is better.

Thus, it is possible to conclude:

1. The metric distance does not appear to be a convenient way for generating weights. The weights obtained by

using metric distance are not always in accordance with the assumption that each observation should be weighted by considering its "proximity to current conditions", Ladd (25, p. 9).

2. Although there is indeterminancy, by using the overall measure of accuracy (TM, TG, TZ and TA), the OLS procedure predicts over the sample period more turning points than either TWR or MWR procedure. Beyond the sample period the OLS only forecast one out of three turning points. In two out of six models the unweighted regression procedure performed better than the weighted ones. TWR and MWR procedures only performed better than the OLS in one model.

For the other two models it was not possible to decide which procedure is superior.

Measures M_{jk} and Z_{jk} are based on arithmetic mean squared errors. On these tests, OLS performed best in two models, TWR in two, and MWR also in two, and TM was smallest for TWR and TZ smallest for MWR.

These results suggest that an analyst who places more weight on criteria based on arithmetic mean squared errors should use a weighted procedure. But results do not clearly identify one weighted method as always superior to the other.

Measures based on absolute errors or geometric mean squared errors give less weight to extreme errors that do

M_{jk} and Z_{jk}. According to A_{jk} and G_{jk}, OLS gives superior results in four of the six models. An analyst who does not want to assign large importance to large errors should choose OLS.

Several suggestions can be made for further research about this subject:

1. To forecast more values for each model, instead of generating two forecasts, it would be convenient to construct a larger set of forecasts for the same model: in total it is possible to generate $t-(k+1)$ forecasts for each model, based on the condition of rank of the matrix of observations $k<t$.

This might lead to unambiguous results about which regression procedure is better.

2. To compare the forecast obtained by the different regression procedures with a method such as the Box-Jenkins method, which is a regression forecasting procedure based on past values.

3. The last metric weight was obtained by adding the two largest metric weights; it would be convenient to assume other values for this last weight and to verify how these values affect the performance of the model for forecasting.

4. To combine the several forecasts in order to obtain a single forecast. Methods developed for this purpose require a larger number of forecasts than are obtained in this thesis.

5. The process of generating forecasts by using weighted regression is cumbersome and expensive and causes difficulties in deriving tests of hypothesis for the estimators.

BIBLIOGRAPHY

1. Ahalt, J. D. and Alvin C. Egbert. "The Demand for Feed Concentrates: A Statistical Analysis." Agricultural Economic Research 17 (April, 1965):41-49.

2. Aitchison, J. and S. D. Silvey. "Maximum Likelihood Estimation of Parameters Subject to Restraints." Analysis of Mathematical Statistics 29 (1958): 813-828.

3. Andrews, D. F., Bickel, P. J., Hempel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. Robust Estimates of Location: Survey and Advances. Princeton, N.J.: Princeton University Press, 1972.

4. Chow, G. C. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." Econometrica 28 (July, 1960):591-606.

5. Chow, G. "A Family of Estimators for Simultaneous Equation Systems." International Economic Review 15 (3, October, 1974):654-666.

6. Dhrymes, P. J., E. Howrey, S. Hymans, J. Kmenta, et al. "Criteria for Evaluation of Econometric Models." Annals of Economic and Social Measurement 1 (July, 1972):291-324.

7. Draper, Norman and H. Smith. Applied Regression Analysis. New York: John Wiley & Sons, 1966.

8. Evans, M. and Edward Green. "The Relative Efficacy of Investment Anticipations." JASA 61 (March, 1966): 104-116.

9. Fair, R. C. "A Comparison of FIML and Robust Estimates of a Nonlinear Econometric Model." NBER Working Paper 15 (October, 1973):206-212.

10. Fisher, F. M. A Priori Information and Time Series Analysis. Amsterdam: North Holland Publishing, Co., 1966.

11. Fisk, P. R. "Models of the Second Kind in Regression Analysis." Journal of the Royal Statistical Society B 29: (1967):266-281.

12. Friedman, Milton. "The Methodology of Positive Economics." In Essays in Positive Economics. Edited by M. Friedman. Chicago: University of Chicago Press, 1966.

13. Friend, Irwin and P. Taubman. "A Short-Term Forecasting Model." Review of Economics and Statistics 6 (August, 1964):229-236.

14. Fuller, Wayne. "Errors in Variables." Lecture Notes. Department of Statistics, Iowa State University, Ames, Iowa, 1977.

15. Granger, C. W. J. "The Typical Spectral Shape of an Economic Variable." Econometrica 34 (January, 1966):150-161.

16. Granger, C. W. J. and P. Newbold. Forecasting Economic Time Series. New York: Academic Press, 1977.

17. Howrey, E. Philip. "Dynamic Properties of a Condensed Version of the Wharton Model." In Econometric Models of Cyclical Behavior. Edited by B. Hickman. Conference on Research in Income and Wealth, Vol. 36, National Bureau of Economic Research, 1972.

18. Howrey, P., L. Klein and M. D. McCarthy. "Notes on Testing the Predictive Performance of Econometric Model." International Economic Review 15 (2, June, 1974):366-383.

19. Huber, P. J. "Robust Statistics: A Review." Annals of Mathematical Statistics 43 (1972):1041-1067.

20. Huber, P. J. "Robust Regression: Asymptotics, Conjectures and Monte Carlo." Annals of Statistics 1 (1973):799-821.

21. Hutchison, T. W. The Significance and Basic Postulates of Economic Theory. London: Macmillan, Co., 1938.

22. Johnston, J. Econometric Methods. New York: McGraw-Hill, Co., 1972.

23. Jorgenson, Dale W., J. Hunter and M. Nadiri. "The Prediction Performance of Econometric Models of Quarterly Investment Behavior." Econometrica 38 (March, 1970):213-224.

24. Klein, Lawrence R. An Essay on the Theory of Economic Prediction. Chicago: Markham Publishing Co., 1970.

25. Ladd, George. "Evaluating Econometric Models and Forecast." ISU-Department of Economics Staff Paper Series 33 (April, 1976):

26. Maddala, G. S. Econometrics. New York: McGraw-Hill Book Co., 1977.

27. Mincer, J. and V. Zarnowitz. "Economic Forecast and Expectation." The Evaluation of Economic Forecast 5th ed. Columbia: Columbia University Press, 1969.

28. Naylor, Thomas H., K. Wertz, and T. Wonnacott. "Spectral Analysis of Data Generated by Simulations Experiments with Econometric Models." Econometrica 37 (April, 1969):333-352.

29. Nelder, J. A. "Regression, Model Building and Invariance," J. Royal Statistical Society A131 (1968):303-329.

30. Pindick, Robert and D. Rubinfeld. Econometric Models and Economic Forecast. New York: McGraw-Hill Book Co., 1976.

31. Platt, Robert B. "Statistical Measures of Forecast Accuracy." In Method and Techniques of Business Forecasting, edited by W. Butter et al. Englewood Cliffs, N.J.: Prentice Hall, 1974.

32. Ramsey, J. B. "Test for Specification Errors in Classical Least Squares Regression Analysis." Journal of the Royal Statistical Society 831 (1969): 350-371.

33. Ryan, M. and M. Abel. "Oats and Barley Acreage Response to Government Programs." Agricultural Economics Research 25 (October, 1973):105-114.

34. Seber, G. A. F. Linear Regression Analysis. New York: John Wiley and Sons, 1977.

35. Shapiro, H. T. "Is Verification Possible? The Evaluation of Large Econometric Models." American Journal of Agricultural Economics 55 (May, 1973): 250-258.

36. Stekler, Herman O. "An Evaluation of Quarterly Judgmental Economic Forecasts." Journal of Business 41 (July, 1968):329-339.

37. Stekler, H. O. "Forecasting with Econometric Models: An Evaluation." Econometrica 36 (July-October, 1968):437-463.

38. Tanzer, Michael D. "Stages and Local Government Debt in the Post-war Period." Review of Economics and Statistics 46 (August, 1964):235-244.

39. Theil, H. A. Applied Economic Forecasting. North Holland: North Holland Publishing Co., 1966.

40. Theil, H. A. Principles of Econometrics. New York: John Wiley & Sons, 1971.

41. Tong, Hun Lee. "The Stock Demand Elasticities of Non-Farm Housing." Review of Economics and Statistics 46 (February, 1964):82-89.

42. Tukey, J. W. "On the Comparative Anatomy of Transformation." Annals of Mathematical Statistics 28 (1957):602-632.

43. Turner, M. E., R. Monroe and H. Lucas. "Generalized Asymptotic Regression and Non-linear Path Analysis." Biometrics 17 (1961):120-143.

44. Zellner, A. An Introduction to Bayesian Inference in Econometrics. New York: John Wiley and Sons, Inc., 1971.