Evaluating Econometric Models And Forecasts

George W. Ladd
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/econ_las_staffpapers

Part of the Econometrics Commons, Economic Theory Commons, Statistical Methodology Commons, Statistical Models Commons, and the Statistical Theory Commons

Recommended Citation
http://lib.dr.iastate.edu/econ_las_staffpapers/150

This Report is brought to you for free and open access by the Economics at Iowa State University Digital Repository. It has been accepted for inclusion in Economic Staff Paper Series by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Evaluating Econometric Models And Forecasts

Abstract
By validation I mean comparing values generated by a model with actual values to determine how well the model reproduces what it was intended to reproduce. Clearly evaluation involves more than this. Evaluation involves judging the appropriateness of the estimation procedures and judging the use made of prior econometric knowledge in constructing the model. A model can be validated by determining its performance during the sample period, or by determining its predictive performance in the post-sample period.

Disciplines
Econometrics | Economic Theory | Statistical Methodology | Statistical Models | Statistical Theory
EVALUATING ECONOMETRIC MODELS AND FORECASTS

by

George W. Ladd

No. 33

April 1976

In this paper I do not propose to try to present complete coverage of the topic assigned to me. For anyone desiring comprehensive coverage of the topic, I recommend starting with the papers presented by Shapiro [14], Fromm [8], Christ [5] and Rausser [13] at the session on The Validity and Verification of Complex Economic Systems held at the Winter 1972 meetings of the Allied Social Science Associations.

Part of my paper presents some purely personal speculations; part presents results of empirical work. The various sections, diverse in nature as they are, all carry the common theme, "Let us use our knowledge of the purpose of our models in estimating and evaluating the models."
INTEGRATION OF ESTIMATION AND VALIDATION CRITERIA

By validation I mean comparing values generated by a model with actual values to determine how well the model reproduces what it was intended to reproduce. Clearly evaluation involves more than this. Evaluation involves judging the appropriateness of the estimation procedures and judging the use made of prior economic knowledge in constructing the model. A model can be validated by determining its performance during the sample period, or by determining its predictive performance in the post-sample period. In testing the goodness of fit of a model we can measure how well the output variables of the model conform with reality in: (a) average values of variables, (b) variations about means of variables, (c) amplitudes of fluctuations over the entire sample, or over specific segments of the sample, (d) number, timing and direction of turning points, and (e) probability distribution of variables. Other could also be listed.

Almost every discussion of model validation will sooner or later discuss turning points: How well does the model predict (or estimate) turning points? No doubt, being able to predict turning points accurately is fun, is of some importance and contributes to one's self-esteem. All the same, it may not be of much scientific significance. In the first place, what is a turning point? Consider the following hypothetical "time series."

90.1, 90.2, 90.3*, 90.2*, 90.3, 90.4, 90.5*, 90.4

Is each number marked by an asterisk a turning point? Is every change in the sign of the first difference a turning point, or is only the first difference with a large absolute value and a change in sign a turning point? Or perhaps a turning point occurs only if the first differences have the same sign for at least $m_0$ months and then have the opposite sign for at least $m_1$ months. Also, a turning point may not represent systematic variation; it may result from random
variation or white noise. Consider equation (1) where $\varepsilon_t$ is random,

$$Y_t = \sum_i \beta_i X_{it} + \varepsilon_t$$

and set $X_{it} = \bar{X}_i$ for all $i$ and $t$. Then $Y_t = \sum_i \beta_i \bar{X}_i + \varepsilon_t$ and

$$\text{Prob} \left[ \text{sign} \left( Y_t - Y_{t-1} \right) \neq \text{sign} \left( Y_{t-1} - Y_{t-2} \right) \right] =$$

$$\text{Prob} \left[ \text{sign} \left( \varepsilon_t - \varepsilon_{t-1} \right) \neq \text{sign} \left( \varepsilon_{t-1} - \varepsilon_{t-2} \right) \right]$$

Any change in the sign of the first difference of $Y_t$ is due to a change in the sign of the first difference of $\varepsilon_t$. It is extremely unlikely that a series of random drawings will yield all $\Delta \varepsilon_t$ of the same sign. In any observed time series we can not know how many of the turning points are due to the systematic influences and how many are due to random influences. I question how scientific it is to develop a model for measurement of systematic influences and then to use an unidentified mixture of systematic and random influences -- the mixture that generates turning points -- as a tool for validation.

In their discussion of validation, many writers seem to give the impression that estimation and validation are completely separate processes. They approach validation by asking, "O.K. You have estimated your model, now how well does it fit reality?" And they suggest several criteria for deciding "how well it fits reality" and commonly none of these criteria were used in selecting an estimation procedure. For example, a person who estimated a model by minimizing sums of squares of deviations of estimated values from actual values will use Theil's inequality coefficient in validation. But an inequality coefficient refers to sum of squares of deviations of estimated first differences about actual first differences. If the ability to predict first differences is the most important criterion, one should select an estimation procedure that minimizes $\sum \Delta \varepsilon_t^2$ (where $\varepsilon_t$ is an estimate of $\varepsilon_t$) rather than one that minimizes $\sum \varepsilon_t^2$. Or, a person will measure the ability of the model to reproduce turning points even though turning points played no role in selection of an estimation procedure. Ideally, if a
criterion is important it should be used in selection of estimates and estimation procedure. I say ideally because one may not be able to develop a computable procedure that will accomplish all his objectives. Suppose one wants unbiased and efficient (or consistent and asymptotically efficient) estimates of coefficients and wants a model that closely reproduces first differences. The second criteria suggests minimization of $\sum \Delta \varepsilon_t^2$. Attainment of both objectives is difficult unless the data possess some special properties. Taking first differences can increase the correlation between successive disturbances. If the independent variables contain serially independent errors of measurement, taking first differences increases the variance of the measurement errors. These two effects of differencing create problems for obtaining consistent and asymptotically efficient estimates.

We may be able to salvage something, however. Suppose it is desired to estimate (1) where $\varepsilon_t = \rho \varepsilon_{t-1} + \mu_t$ where $\mu_t$ is random. Then $\Delta \varepsilon_t = (\rho-1)\varepsilon_{t-1} + \mu_t$ and

$$
(2) \quad \Delta Y_t - (\rho-1)Y_{t-1} = \sum_i \beta_i (\Delta X_{it} - (\rho-1)X_{it-1}) + \varepsilon_t
$$

If $\rho$ were known, minimizing $\sum \varepsilon_t^2$ would minimize the residual sum of squares of first differences. But $\rho$ is typically not known. If we are willing to relax our objective of minimizing residual sum of squares of first differences in $Y$ to allow minimization of residual sum of squares of $\Delta Y_t - (\rho-1)Y_{t-1}$, we can apply a nonlinear estimation procedure to (2) to obtain consistent, and perhaps asymptotically efficient, estimates of the $\beta_i$ and $\rho$.

We may also be able to accomplish something by judicious use of restrictions especially if we are willing to replace unbiasedness by minimum mean square error. (See Appendix I for a brief discussion of restricted regression.) For example, suppose errors are temporally independent so use of first differences in the linear statistical model will lead to inefficient (and possibly biased)
estimates but we do want to exercise some control over errors in estimated first differences. We might then use restricted regression to minimize $\varepsilon e_t^2$ while imposing the two restrictions $\Delta \bar{Y}(1) = \sum_{i} \beta_i \Delta \bar{X}_i(1)$ and $\Delta \bar{Y}(2) = \sum_{i} \beta_i \Delta \bar{X}_i(2)$ where $\Delta \bar{Y}(1)$ and $\Delta \bar{Y}(2)$ are mean values of $\Delta Y_t$ in the first and second halves of the sample period, and $\Delta \bar{X}_i(1)$ and $\Delta \bar{X}_i(2)$ are the means of $\Delta X_{it}$ in the two halves of the sample period.

A person who believes it important to predict turning points could impose constraints requiring that the estimated function exactly reproduce all turning points in the sample period. Suppose, for example, that $\Delta Y_q$ and $\Delta Y_r$ represent turning points. He can minimize $\varepsilon e_t^2$ subject to the requirements that

$$
\begin{align*}
\Delta Y_q &= \sum_{i} \beta_i \Delta X_{iq} \\
\Delta Y_r &= \sum_{i} \beta_i \Delta X_{ir}
\end{align*}
$$

Imposition of (3) introduces two problems, or perhaps two symptoms of the same problem. One problem is as follows. Setting $t = q$ and $t = r$ in (1) and first differencing yields

$$
\begin{align*}
\Delta Y_q &= \sum_{i} \beta_i \Delta X_{iq} + \varepsilon_q - \varepsilon_{q-1} \\
\Delta Y_r &= \sum_{i} \beta_i \Delta X_{ir} + \varepsilon_r - \varepsilon_{r-1}
\end{align*}
$$

Then, (3) implies (4)

$$
\begin{align*}
\varepsilon_q &= \varepsilon_{q-1} \text{ and } \varepsilon_r = \varepsilon_{r-1}
\end{align*}
$$

Let $U_{ij}$ represent an $n \times n$ matrix having a unit entry in row $i$ and column $j$ and zeroes everywhere else. Then (4) implies

$$
\begin{align*}
E(\varepsilon \varepsilon') &= (I + U_q, q-1 + U_q, q + U_r, r-1 + U_r, r)\sigma^2
\end{align*}
$$

The standard results on restricted regression, summarized in Appendix I, depend upon the assumption $E(\varepsilon \varepsilon') = I\sigma^2$. Clearly, if (3) is to be assumed, (5) should be assumed in obtaining the restricted estimates and in deriving variances, mean square errors and hypothesis tests.

The second problem arises also from the nature of (3). In Appendix I, the
restrictions are written $R' \beta = r$ where $R'$ and $r$ are assumed known constant matrices. If turning points are stochastic events -- influenced by $\Delta c_t$ -- then, using turning points to determine $R$ and $r$, as is done in (3), violates the assumption of fixed $R$ and $r$ because $R$ and $r$ are functions of random variables. Again, the results in Appendix I do not apply. It may be that either, or both, of these difficulties can be easily handled by using the restricted maximum likelihood procedures of Aitchison and Silvey \[1, 2, 15\]. See Byron \[4\] and Dhrymes, et. al. \[6\] for summaries of their work.

Of course, I suppose, one can argue that a person does not believe turning points are systematic and predictable unless he believes they are entirely determined by elements of $X$. And he would be quite willing to assume that $R'$ and $r$ specified in (3) are fixed because they are determined by known exogenous variables.

One criterion that has been suggested for validation is a comparison of actual and estimated amplitudes of fluctuation over a specific segment of the sample. This, and the criterion of reproducing turning points, suggest to me that it may be more important to obtain a good fit to some sample observations than to other sample observations. In other words, some observations may be more important than others. If some observations are more important than others, why do we ignore this in estimation and wait until we reach the validation stage to bring it up? It seems to me it would make sense to use this information in estimation by assigning different weights to different observations. Classical least squares estimates are obtained by minimizing $\Sigma e_t^2$. What about minimizing a statistical loss function, or a disutility function obtained by assigning different weights to different squared residuals, that is, minimize $\Sigma p_t e_t^2$ where each $p_t$ is a positive known constant. One simple possibility would be to assign one set of weights to the earliest observations in the sample period and
to assign larger weights to the later observations. Suppose for example one were going to use \( p_1 \sum_{t \leq n/2} e_t^2 + p_2 \sum_{t > n/2} e_t^2 \). To assign values to \( p_1 \) and \( p_2 \) consider the total differential of \( e'Pe \).

\[
de'e'Pe = \sum_t p_t de_t^2
\]

and suppose only \( e_1 \) and \( e_n \) are to vary and the total differential is to equal zero.

\[
p_1 de_1^2 + p_2 de_n^2 = 0; \quad p_2/p_1 = -de_1^2/de_n^2
\]

If \( e_n^2 \) rises by \( de_n^2 \), by how much must \( e_1^2 \) change in order for you to be exactly as satisfied with the new -- after the changes -- weighted sum of squares as you were before the change? Suppose you feel that if \( e_n^2 \) rises by one unit \( (de_n^2 = 1) \) you will need to reduce \( e_1 \) by 2 units \( (de_1^2 = -2) \) in order to be equally satisfied with the new as with the original situation. Then \( p_2/p_1 = -(2/1) = 2 \) or \( p_2 = 2p_1 \). Setting \( p_1 = 1 \), then, \( p_2 = 2 \).

The \( p_i \) are marginal disutilities, the \( p_i/p_d \) are marginal rates of substitution. The estimation procedure that will minimize the loss or disutility function is to be used.

Suppose \( \beta \) is estimated by minimizing \( e'Pe \) where \( P \) is a diagonal matrix whose \( t \)-th element is \( p_t > 0 \) and each \( p_t \) is a fixed number selected independently of \( e \). \( e'Pe \) is minimized by the unbiased estimate \( \hat{\beta}_p \)

\[
\hat{\beta}_p = (X'PX)^{-1}X'PY
\]

and the covariance matrix of \( \hat{\beta}_p \) is

\[
D(\hat{\beta}_p) = (X'PX)^{-1}X'^2X(X'PX)^{-1} \sigma^2
\]

Suppose \( \hat{\beta}_p \) is used in forecasting \( Y_f \) and write the forecast as \( Y_{fp} = X_f'\hat{\beta}_p \).

\( Y_{fp} \) is an unbiased forecast and its variance is

\[
D(Y_{fp}) = X_f'D(\beta_p)X_f + \sigma^2
\]
Given an X matrix, or a set of typical X matrices, does there exist a P such that \( D(\beta_p) < D(\hat{\beta}) \) or \( D(Y_{fp}) < D(\hat{Y}_f) \) where \( \hat{\beta} \) and \( \hat{Y}_f \) are the classical least squares estimator and forecast.

The requirement that each \( p_t \) be selected independently of \( \epsilon_t \) raises the same question discussed earlier if large values of \( p_t \) are assigned to turning points: the \( p_t \) are functions of the \( \epsilon_t \).

Before going farther into the question of determining weights, I want to look into something I call "the ubiquity of structural change" for the light it may cast on the topic. Most of the prior economic knowledge we draw on in constructing econometric models is provided by comparative static analysis of constrained optimization problems. These comparative static analyses have one implication we have overlooked in our econometric work, and this implication may provide some insight into the issue of weighting observations. The general point to be made will be illustrated by analysis of a competitive, single-output, multiple-input firm. Let \( p_0 \) and \( q_0 \) be price and quantity of output, and let \( p_j \) and \( q_j \) be price and quantity of the j-th input, and let the firm's production function be \( q_0 = f(q_1, q_2, \ldots, q_m) \). The firm's profit is \( \pi = p_0q_0 - \sum_{j=1}^{m} p_j q_j \).

The appropriate Lagrangean for profit maximization subject to the production function is

\[
L = p_0q_0 - \sum_{j=1}^{m} p_j q_j - \lambda[q_0 - f(q_1, q_2, \ldots, q_m)]
\]

The first-order conditions are, letting \( f_i = \partial f / \partial q_i \),

\[
\begin{align*}
\partial L / \partial \lambda &= f - q_0 = 0 \\
\partial L / \partial q_0 &= p_0 - \lambda = 0 \\
\partial L / \partial q_i &= -p_i + \lambda f_i \quad i = 1, 2, \ldots, m
\end{align*}
\]

From these we can derive, among other relations,

\[
\begin{align*}
(6) \quad p_0 &= \lambda \\
(7) \quad p_i / f_i &= \lambda
\end{align*}
\]

Suppose product price and all input prices vary. What happens to the profit-
maximizing level of output? Let $dp_0$ be the variation in output price and $dp_i$ be the variation in the $i$-th input price. Then the resulting variation in the profit-maximizing level of output -- the supply function -- is

$$dq_0 = \left(\frac{\partial q_0}{\partial p_0}\right) dp_0 + \sum_{i=1}^{m} \left(\frac{\partial q_0}{\partial p_i}\right) dp_i$$

We need the bordered Hessian of the Lagrangean. This bordered Hessian is, letting $f_{ij} = \frac{\partial^2 f}{\partial q_i \partial q_j}$

$$H = \begin{pmatrix} 0 & -1 & f_1 & f_2 & \cdots & f_m \\ -1 & 0 & 0 & 0 & \cdots & 0 \\ f_1 & 0 & \lambda f_{11} & \lambda f_{12} & \cdots & \lambda f_{1m} \\ f_2 & 0 & \lambda f_{21} & \lambda f_{22} & \cdots & \lambda f_{2m} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ f_m & 0 & \lambda f_{m1} & \lambda f_{m2} & \cdots & \lambda f_{mm} \end{pmatrix}$$

Let $e_j$ be the $j$-th unit (column) vector and define the matrices $H_j$ = matrix obtained from $H$ by replacing $j$-th column of $H$ by $-e_j$.

Then

$$dq_0 = \left(\frac{\det H_2}{\det H}\right) dp_0 + \sum_{i=3}^{m+2} \left(\frac{\det H_i}{\det H}\right) dp_i$$

Now notice that by using (6) and (7), $H$ can be written as

$$H = \begin{pmatrix} 0 & -1 & p_j/p_0 \\ -1 & 0 & 0 \\ p_j/p_0 & 0 & p_0 f_{ij} \end{pmatrix}$$

Thus, $\det H$ is a function of $p_0, p_1, p_2, \ldots, p_m$, as is each $\det H_i$. Denote $\frac{\det H_i}{\det H}$ by $R_i(P)$. $R_i$ to denote ratio and $P$ to denote the fact that the numerator and denominator, and hence the ratio, are functions of the vector of prices $P$. Then we can write

$$(8) dq_0 = R_2(P) dp_0 + \sum_{i>2} R_i(P) dp_i = \left(\frac{\partial q_0}{\partial p_0}\right) dp_0 + \sum_{i>2} \left(\frac{\partial q_0}{\partial p_i}\right) dp_i$$
The values of $\partial q_0/\partial p_j$ are functions of existing prices. To put this into an estimation and forecasting context, let $p_{0t}$, $p_{it}$ and $q_{0t}$ be the values of $p_0$, $p_i$ and $q_0$ in the $t$-th period in a time series sample, $(t = 1, 2, \ldots, n)$. For periods $t$ and $t+1$, (8) can be written

$$q_{ot+1} - q_{0t} = R_2(P_t)(p_{0t+1} - p_{0t}) + \Sigma R_i(P_t)(p_{it+1} - p_{it})$$

where $R_i(P_t)$ denotes a function of prices at time $t$. For forecasting the value of $q_{on+1}$,

$$q_{on+1} - q_{0n} = R_2(P_n)(p_{0n+1} - p_{0n}) + \Sigma R_i(P_n)(p_{in+1} - p_{in})$$

The $R_2(P_n)$ and the $R_i(P_n)$ are the coefficients to be estimated for use in making the forecast. I do not call the coefficients "parameters" because their values are functions of current prices. The effect of a unit change in any one price on $q_0$, therefore, depends upon the values of that price and of all other prices. This is indicated by the presence of the time subscript in $R_i(P_t)$ and $R_i(P_n)$.

The presence of the time subscript indicates $\partial q_0/\partial p_i$ is not constant for all sample observations. This might suggest any one of three different possibilities: (A) the use of a random coefficients model, (B) a nonlinear function to allow each partial derivative to be a function of prices, or (C) use of weights. The first possibility does not appeal to me for two reasons. (1) Prices are serially correlated over time and I would therefore expect each partial derivative to be serially correlated. (2) A random coefficient model can be expressed $Y_t = \Sigma \beta_i x_{it} + \varepsilon_t$, $\beta_{it} = \bar{\beta}_i + \mu_{it}$, $\beta_i$ a parameter. In the present problem $\bar{\beta}_i$ is a function of sample values of prices. Changing the sample will change $\bar{\beta}_i$.

As shown in their derivation, each coefficient is a function of "current conditions." If we are in period $n$ and want to make forecasts for $n+1$, the "current conditions" of period $n$ are the most important conditions to us. Sample periods in which 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. To estimate $R_2 \left( P_n \right)$ and the $R_1 \left( P_n \right)$ we might weight each observation in accordance with its "proximity to current conditions."

Two possible measures of proximity are temporal distance and metric distance. To use temporal distance as weights, multiply each observation by the reciprocal of the square root of its distance in time from period $n+1$. Observation $t$ would be weighted by multiplication by $1/(n+1-t)^{1/2}$. Let $X_t$ by the (column) vector of independent variables in the $t$-th sample period. The metric distance between $X_t$ and $X_n$ is

$$d(t,n) = [(X_t - X_n)'(X_t - X_n)]^{1/2} = \sum_i (X_{it} - X_{in})^2$$

For $t < n$, weight the $t$-th observation by $1/d(t,n)$. The weight for period $n$ will have to be arbitrary. It should, however, satisfy

$$w_n \geq \max_{t} (1/d(t,n))$$

Using either metric distance or temporal distance to determine weights is consistent with the assumption that the $p_t$ are independent of the errors.

The motivation behind this section has been the idea that if a criterion is sufficiently important to use in validating a model, it is sufficiently important to be incorporated into the estimation procedure. Three possibilities have been suggested: (a) If accurate reproduction of first differences is important, develop appropriate estimation procedures that take account of effect of first differencing on autocorrelation properties of the disturbances. (b) Use restricted regression. (c) Use weighted estimation procedures to assign more importance to some observations than to others.

MULTIPLE OUTPUT PROBLEM

Suppose, to take a simple case, we are concerned with a model that has five endogenous variables and we have three estimated versions of each structural
equation. The different versions may represent different estimation procedures or may differ in exogenous variables. We then have 243 (=3^5) versions of our five equation model.

Most validation procedures are multi-step one-variables-at-a-time-procedures. They help to answer the question: Which one of the 243 versions of my model does the best job of simulating or estimating endogenous variable V_i? These procedures give less help in answering the question: Which version of my model does the best job of simulating all the endogenous variables?

Suppose it happened that the best predictions of our five endogenous variables were provided by five different versions of the model. Each variable requires a different version of the model for the best prediction. Resource limitations may require us to pick one version for future work. Which version of our model is the superior version? To answer this question, I go back to the theme of this paper: "Let us use our knowledge of the purpose of our models in estimation and evaluation." Develop a weighted criterion consistent with the purpose of the model. Suppose we have agreed on some statistic that we will use to measure the performance of each individual version of an equation or the goodness of fit of each variable. Let S_{im} be the value of the statistic for the i-th variable in version m of our model. Assign a weight p_i to each variable and compute the weighted statistic for version m

\[ PS_m = \sum_{i} p_i S_{im} \]

If a small value of S_{im} is desirable, the model yielding the smallest value of PS_m is the preferred model. If a large value of S_{im} is desirable, the model yielding the largest value of PS_m is the preferred model. E'Pe of the previous section is a disutility or statistical loss function that assigns more weight to certain observations on a variable than to other observations on the same variable. PS_m is a disutility (or utility) function that assigns more importance
to some variables than to others. One might base the weights on marginal rates of substitution of utility, as was done previously.

We used a different procedure in developing a system of weights for the model Alan Rahn [12] constructed in his thesis. The prime objective of his study was to develop a model for prediction of farm prices and farm marketings of cattle, hogs, sheep, broiler chickens and turkeys. The model contained five subsectors, one for each of these five commodities. Rahn and Gene Futrell developed a weighting system for each subsector and then an over-all weighting system. In each subsector they arbitrarily assigned a "priority index" value of 1,000 to the commercial production variable. Using this as a basis they assigned a priority index to each of the other variables in the sector. The priority indexes represented their combined judgment as to the relative importance of the various variables. The priority indexes were then converted to weights with the sum of the weights in each subsector equalling one. Tables 1 through 5 present the priority indexes and the weights for each subsector. These weights were used to compare various versions of each subsector model.

To obtain a set of over-all weights for use in validating the complete model, each subsector was assigned a weight representing the percentage of farm income from the five commodities contributed by the farm product in the respective subsector during 1967-71. These proportions are presented in Table 6.

In comparing, for example, different versions of the beef subsector, each variable in Table 1 was assigned the weight listed in that table. In comparing different versions of the entire model, the weight assigned to each variable in the beef subsector was obtained by multiplying that variable's weight listed in Table 1 by 0.6635, from Table 6.

What about incorporating into systems estimation procedures the idea of weighting different endogenous variables according to their relative importance?
Table 1. Priority Indexes and Weights Used in Constructing Beef Subsector Validation Index

<table>
<thead>
<tr>
<th>Variable</th>
<th>Priority Index</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Wholesale beef price</td>
<td>2,000</td>
<td>0.1581</td>
</tr>
<tr>
<td>2. Slaughter steer price</td>
<td>1,800</td>
<td>0.1423</td>
</tr>
<tr>
<td>3. January 1 inventory of beef cows and heifers that have calved</td>
<td>1,700</td>
<td>0.1344</td>
</tr>
<tr>
<td>4. January 1 inventory of steers 500 pounds and over</td>
<td>1,600</td>
<td>0.1265</td>
</tr>
<tr>
<td>5. January 1 inventory of heifers 500 pounds and over not being kept for milk cow replacement</td>
<td>1,200</td>
<td>0.0949</td>
</tr>
<tr>
<td>6. January 1 inventory of heifers, steers and bulls under 500 pounds</td>
<td>1,100</td>
<td>0.0870</td>
</tr>
<tr>
<td>7. Commercial beef production</td>
<td>1,000</td>
<td>0.0790</td>
</tr>
<tr>
<td>8. Commercial cattle slaughter</td>
<td>900</td>
<td>0.0711</td>
</tr>
<tr>
<td>9. Average live weight of commercial cattle slaughter</td>
<td>300</td>
<td>0.0237</td>
</tr>
<tr>
<td>10. Feeder calf price</td>
<td>250</td>
<td>0.0198</td>
</tr>
<tr>
<td>11. Net foreign trade in beef</td>
<td>225</td>
<td>0.0178</td>
</tr>
<tr>
<td>12. Commercial steer slaughter</td>
<td>200</td>
<td>0.0158</td>
</tr>
<tr>
<td>13. Commercial heifer slaughter</td>
<td>150</td>
<td>0.0118</td>
</tr>
<tr>
<td>14. Commercial slaughter of beef and milk cows</td>
<td>100</td>
<td>0.0079</td>
</tr>
<tr>
<td>15. Cold storage holdings of beef</td>
<td>75</td>
<td>0.0059</td>
</tr>
<tr>
<td>16. Dressing yield, commercial cattle slaughter</td>
<td>50</td>
<td>0.0040</td>
</tr>
<tr>
<td>17. Commercial civilian beef consumption</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Source: Rahn, [12, p. 154].
### Table 2. Priority Indexes and Weights Used in Constructing Pork Subsector Validation Index

<table>
<thead>
<tr>
<th>Variable</th>
<th>Priority</th>
<th>Proportional weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wholesale pork price</td>
<td>2000</td>
<td>0.2524</td>
</tr>
<tr>
<td>Price of barrows and gilts</td>
<td>1800</td>
<td>0.2271</td>
</tr>
<tr>
<td>Sows farrowing</td>
<td>1500</td>
<td>0.1893</td>
</tr>
<tr>
<td>Commercial pork production</td>
<td>1000</td>
<td>0.1262</td>
</tr>
<tr>
<td>Commercial hog slaughter</td>
<td>900</td>
<td>0.1136</td>
</tr>
<tr>
<td>Average live weight of slaughter hogs</td>
<td>300</td>
<td>0.0378</td>
</tr>
<tr>
<td>Commercial slaughter of barrows and gilts</td>
<td>200</td>
<td>0.0252</td>
</tr>
<tr>
<td>Commercial sow slaughter</td>
<td>125</td>
<td>0.01582</td>
</tr>
<tr>
<td>Cold storage holdings of pork</td>
<td>100</td>
<td>0.0126</td>
</tr>
<tr>
<td>Commercial civilian pork consumption</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Source: Rahn, [12, p. 155].

### Table 3. Priority Indexes and Weights Used in Constructing Lamb Subsector Validation Index

<table>
<thead>
<tr>
<th>Variable</th>
<th>Priority</th>
<th>Proportional weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wholesale lamb price</td>
<td>2400</td>
<td>0.3211</td>
</tr>
<tr>
<td>Price of sheep and lambs</td>
<td>2200</td>
<td>0.2943</td>
</tr>
<tr>
<td>Ewes one year and over on farms</td>
<td>1800</td>
<td>0.2408</td>
</tr>
<tr>
<td>Commercial lamb and mutton production</td>
<td>1000</td>
<td>0.1338</td>
</tr>
<tr>
<td>Cold storage holdings of lamb and mutton</td>
<td>75</td>
<td>0.0100</td>
</tr>
<tr>
<td>Commercial civilian lamb and mutton consumption</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Source: Rahn, [12, p. 155].
Table 4. Priority Indexes and Weights Used in Constructing Broiler Subsector Validation Index

<table>
<thead>
<tr>
<th>Variable</th>
<th>Priority index</th>
<th>Proportional weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Wholesale broiler price</td>
<td>2000</td>
<td>0.3670</td>
</tr>
<tr>
<td>2. Broiler type chick hatchings</td>
<td>1500</td>
<td>0.2752</td>
</tr>
<tr>
<td>3. Commercial broiler meat production</td>
<td>1000</td>
<td>0.1835</td>
</tr>
<tr>
<td>4. Broiler chicks tested for pullorum</td>
<td>800</td>
<td>0.1468</td>
</tr>
<tr>
<td>5. Cold storage broiler holdings</td>
<td>150</td>
<td>0.0275</td>
</tr>
<tr>
<td>6. Commercial civilian consumption of broiler meat</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Source: Rahn, [12, p. 156].

Table 5. Priority Indexes and Weights Used in Constructing Turkey Subsector Validation Index

<table>
<thead>
<tr>
<th>Variable</th>
<th>Priority index</th>
<th>Proportional weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Wholesale turkey price</td>
<td>2000</td>
<td>0.3077</td>
</tr>
<tr>
<td>2. Heavy breed turkey hatch</td>
<td>1500</td>
<td>0.2308</td>
</tr>
<tr>
<td>3. Commercial turkey meat production</td>
<td>1000</td>
<td>0.1538</td>
</tr>
<tr>
<td>4. Cold storage holdings of turkey meat</td>
<td>900</td>
<td>0.1385</td>
</tr>
<tr>
<td>5. Heavy breed turkeys tested for pullorum</td>
<td>700</td>
<td>0.1077</td>
</tr>
<tr>
<td>6. Light breed turkey hatch</td>
<td>300</td>
<td>0.0462</td>
</tr>
<tr>
<td>7. Light breed turkeys tested for pullorum</td>
<td>100</td>
<td>0.0153</td>
</tr>
<tr>
<td>8. Commercial civilian consumption of turkey meat</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Source: Rahn, [12, p. 156].
Table 6. Weights Used in Constructing Aggregate Validation Index

<table>
<thead>
<tr>
<th>Subsector</th>
<th>Cash receipts(^a) (1967-1971 average)</th>
<th>Proportional weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beef</td>
<td>12,803,447</td>
<td>0.6635</td>
</tr>
<tr>
<td>Pork</td>
<td>4,285,753</td>
<td>0.2224</td>
</tr>
<tr>
<td>Lamb</td>
<td>322,935</td>
<td>0.0168</td>
</tr>
<tr>
<td>Broiler</td>
<td>1,409,075</td>
<td>0.0732</td>
</tr>
<tr>
<td>Turkey</td>
<td>465,195</td>
<td>0.0241</td>
</tr>
<tr>
<td>All subsectors</td>
<td>19,286,405</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

\(^a\) Cash receipts from farm marketings and value of products consumed in farm households.

Source: Rahn, [12, p. 157].

This can easily be done in the case of a system of equations each of which contains only one endogenous variable. This is the kind of system Zellner dealt with in his "seemingly unrelated regressions." Suppose we have a set of \(G\) equations, each containing only one endogenous variable, and the \(i\)-th equation in the system can be written \(y_i = \beta_1 x_i + \mu_i\). Zellner's procedure estimates all of the \(\beta_1, \beta_2, ..., \beta_G\) simultaneously. His procedure can be obtained by minimizing a sum of squares of transformed errors. The transformation is performed to make the errors independent and homoscedastic. Zellner's estimates are exactly equal to the least squares estimates obtained by applying least squares to each equation separately if either:

(A) \(Z_1 = Z_2 = ... = Z_G\), or

(B) Errors in different equations are independent.

If neither of these conditions is satisfied, Zellner's estimates are more efficient than the least squares estimates of separate equations.
Weights can be assigned to the $G$ different endogenous variables, and weighted least squares estimates can be obtained by minimizing a weighted sum of squares of transformed errors. The weighted estimates will differ from those obtained by least squares estimation of each equation independently even though conditions (A) or (B) preceding are satisfied. Appendix II presents a derivation of these weighted estimates.

Commonly, of course, in the systems of equations we are concerned with, some equations contain more than one endogenous variable. One method of estimating such systems of equations is three-stage-least-squares (3SLS). From a mathematical standpoint, 3SLS estimates can be derived by minimizing a sum of squares of the transformed errors from the structural equations. It is, therefore, mathematically possible to derive weighted 3SLS estimates by minimizing a weighted sum of squares of the transformed errors. The weighted sum of squares of the transformed errors is such an odd-looking function that I am not sure it has any meaning, though.

Many economists have recently recognized that the United States agricultural economy is not a closed system and we can no longer adequately understand developments in our agricultural sector unless we recognize the interdependencies among nations. As Karl Fox wrote in his 1973 report [7, pp. 3, 4] --

"I see no evidence that any U.S. agency did an adequate job of forecasting economic developments during 1972-73. To the best of my knowledge, no U.S. agency has an adequate model of the world economy, or even an adequate conceptual framework within which to discuss interactions among the food, agricultural and other sectors of the world economy. In brief, I believe the whole Federal establishment is ill-prepared in terms of data, models, analytical procedures, and patterns of interagency communication for the tasks of forecasting and policy formation in the "open" economy of 1973."
"The problem is to communicate about agricultural and other developments on a world-wide basis. It is foolish to regard the problem of agricultural intelligence as separable from that of intelligence about the world economy as a whole, and it is foolish not to have, and use, models of the world economy which incorporate agriculture as one of a number of interacting sectors."

In this section I will briefly report on part of a thesis recently completed at Iowa State University by Dyaa K. Abdou [11] in which United States foreign trade in beef was treated as endogenous. The starting point for Abdou's study was a re-estimated version of the Rahn model I referred to earlier. Whereas, in Rahn's model, beef is treated as a homogenous product, in Abdou's model, beef is disaggregated into fed beef and nonfed beef, and United States import of nonfed beef is endogenous. Define

\[
\begin{align*}
\text{NEXSA}(L) &= \text{South America's current net exports of beef and veal}, \\
\text{NEXOC}(L) &= \text{Oceania's current net exports of beef and veal}, \\
\text{NIMWE}(L) &= \text{Western Europe's current net imports of beef and veal}, \\
\text{IMUS}(L) &= \text{United States' current beef and veal imports}, \\
\text{NIMRW}(L) &= \text{Rest of world's current net imports of beef and veal}, \\
\text{BQSA} &= \text{South America's total current beef and veal production}, \\
\text{T}(L) &= \text{linear time trend}, \\
\text{BQOC}(L) &= \text{Oceania's current total beef and veal production}, \\
\text{CEOCL} &= \text{Current per capita private final consumption expenditure in South America, in United States dollars}, \\
\text{BQWE}(L) &= \text{Western Europe's current total beef and veal production}, \\
\text{CEWE}(L) &= \text{Current per capita private final consumption expenditure for Western Europe, in United States dollars}, \\
\text{NFBPW}(L-1) &= \text{Wholesale utility cow price per cwt., previous period}, \\
\text{BQ}(L-1) &= \text{United States' total commercial beef and veal production, previous period, and}
\end{align*}
\]
DYN(L) = Current per capita disposable personal income in the United States. Variables on production and foreign trade were measured in carcass weight equivalents. The preceding variables are all measured on an annual basis.

The estimated structural equations (obtained by 3SLS) in the world trade subsector were [11, pp. 78-80]: (directly under each coefficient, in parentheses, is the ratio of the coefficient to its standard error)

\[
\begin{align*}
(10) \quad \text{NEXSA}(L) &= -3370.0 + 0.4006 \times \text{NEXOC}(L) + 0.3558 \times \text{NIMWE}(L) \\
&\quad -0.3272 \times \text{IMUS}(L) - 0.0327 \times \text{NIMRW}(L) + 0.4286 \\ 
&\quad \times \text{BQSA}(L) - 60.55 \times T(L) \\
&\quad \text{(1.326)}
\end{align*}
\]

\[
\begin{align*}
(11) \quad \text{NEXOC}(L) &= -787.4 - 0.1578 \times \text{NEXSA} - 0.0647 \times \text{NIMWE}(L) + 0.3401 \\
&\quad \times \text{IMUS}(L) + 0.6101 \times \text{NIMRW}(L) + 0.6340 \times \text{BQOC}(L) + 0.2721 \\ 
&\quad \times \text{CEOC}(L) - 31.6500 \times T(L) \\
&\quad \text{(1.336)}
\end{align*}
\]

\[
\begin{align*}
(12) \quad \text{NIMWE}(L) &= 429.0 + 0.5532 \times \text{NEXSA}(L) + 0.4431 \times \text{NEXOC}(L) + 0.1578 \\
&\quad \times \text{IMUS}(L) + 0.4691 \times \text{NIMRW}(L) - 0.3835 \times \text{BQWE}(L) + 0.0541 \\ 
&\quad \times \text{CEWE}(L) + 71.47 \times T(L) \\
&\quad \text{(1.625)}
\end{align*}
\]

\[
\begin{align*}
(13) \quad \text{IMUS}(L) &= -4338.0 + 1.363 \times \text{NEXSA}(L) + 1.293 \times \text{NEXOC}(L) - 1.401 \\
&\quad \times \text{NIMWE}(L) + 0.2243 \times \text{NIMRW}(L) - 34.5200 \times \text{NFBPW}(L-1) + 0.1885 \\ 
&\quad \text{(2.523)}
\end{align*}
\]
\[ BQ(L-1) + 1.526 \text{ DYN}(L) - 388.7 \text{ T}(L) \]

\[(1.207)\]

(14) \[ \text{NEXSA}(L) + \text{NEXOC}(L) - \text{NIMWE}(L) - \text{IMUS}(L) - \text{NIMRW}(L) = 0 \]

The reduced form equations derived from this system were used in prediction and simulation and in obtaining elasticities for U.S. imports with respect to foreign regions' production. The reduced form equation for \( \text{IMUS}(L) \) was the equation of concern in Kamal-Abdou's study.

The yearly level of imports for the U.S., \( \text{IMUS}(L) \), was transformed to quarterly imports according to the following equation.

(15) \[ \text{IMUS}(I) = g \cdot (\text{IMUS}(L)) \]

where \( g = \)

\[ \begin{align*}
   &0.221 \quad \text{for } I = 1 \\
   &0.225 \quad \text{for } I = 2 \\
   &0.300 \quad \text{for } I = 3 \\
   &0.254 \quad \text{for } I = 4
\end{align*} \]

Those quarterly import levels were than used in obtaining the nonfed beef per capita civilian consumption, \( \text{NFBCN}(I) \), through the following identity.

(16) \[ \text{NFBCN}(I) = \frac{[\text{NFBQ}(I) - .5 \text{ MBC}(I) + \text{IMUS}(I)]}{\text{P}(I)} \]

\( \text{MBC}(I) = \) U.S. military consumption of commercial beef, current quarter, \( \text{NFBQ}(I) = \) commercial U.S. production of nonfed beef, current quarter, and \( \text{P}(I) = \) United States civilian resident population, current quarter.

Table 7 presents the priority indexes and weights used in Abdou's study in validation of various versions of the fed and nonfed beef subsector model. For other subsectors, Abdou used the weights in Tables 2 through 6. For the \( S_{im} \) in expression (9) he used an average percentage error and Theil's inequality coefficient. Results obtained from the sample period are summarized in Table 8. SIMU VI is Abdou's model and SIMU V is a model in which beef is treated as a homogeneous product. According to the Theil's inequality criteria, partitioning
Table 7. Priority Indexes and Weights Used in Constructing Beef Subsector Validation Index

<table>
<thead>
<tr>
<th>Variable</th>
<th>Priority index</th>
<th>Proportional weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Wholesale fed beef price</td>
<td>1334</td>
<td>0.0963</td>
</tr>
<tr>
<td>2. Slaughter steer price</td>
<td>1250</td>
<td>0.0903</td>
</tr>
<tr>
<td>3. January 1 inventory of beef cows and heifers that have calved</td>
<td>1240</td>
<td>0.0895</td>
</tr>
<tr>
<td>4. January 1 inventory of steers 500 pounds and over</td>
<td>1200</td>
<td>0.0866</td>
</tr>
<tr>
<td>5. January 1 inventory of heifers, steers and bulls under 500 pounds</td>
<td>985</td>
<td>0.0711</td>
</tr>
<tr>
<td>6. Commercial fed beef production</td>
<td>700</td>
<td>0.0505</td>
</tr>
<tr>
<td>7. January 1 inventory of heifers 500 pounds and over</td>
<td>680</td>
<td>0.0491</td>
</tr>
<tr>
<td>8. Wholesale nonfed beef price</td>
<td>666</td>
<td>0.0481</td>
</tr>
<tr>
<td>9. Cattle and calf placements on feed</td>
<td>650</td>
<td>0.0469</td>
</tr>
<tr>
<td>10. Calf crop</td>
<td>640</td>
<td>0.0462</td>
</tr>
<tr>
<td>11. Price of slaughter utility cows</td>
<td>600</td>
<td>0.0433</td>
</tr>
<tr>
<td>12. January 1 inventory of milk cows and heifers that have calved</td>
<td>580</td>
<td>0.0419</td>
</tr>
<tr>
<td>13. Fed cattle marketed, 23 major states</td>
<td>550</td>
<td>0.0397</td>
</tr>
<tr>
<td>14. January 1 inventory of heifers 500 pounds and over being kept for beef cow replacements</td>
<td>490</td>
<td>0.0354</td>
</tr>
<tr>
<td>15. January 1 inventory of heifers 500 pounds and over being kept for milk cow replacements</td>
<td>470</td>
<td>0.0339</td>
</tr>
<tr>
<td>16. January 1 inventory of bulls 500 pounds and over</td>
<td>465</td>
<td>0.0336</td>
</tr>
<tr>
<td>17. Total nonfed cattle and calves marketed</td>
<td>450</td>
<td>0.0325</td>
</tr>
<tr>
<td>18. Feeder calf price</td>
<td>350</td>
<td>0.0253</td>
</tr>
<tr>
<td>19. Commercial nonfed beef production</td>
<td>300</td>
<td>0.0217</td>
</tr>
<tr>
<td>20. Average dressing weight, fed cattle</td>
<td>120</td>
<td>0.0087</td>
</tr>
<tr>
<td>21. U.S. beef and veal imports</td>
<td>75</td>
<td>0.0054</td>
</tr>
<tr>
<td>22. Average dressing weight for nonfed cattle and calves</td>
<td>55</td>
<td>0.0040</td>
</tr>
</tbody>
</table>

Source: Abdou, [11, p. 106].
beef into fed and nonfed improved not only the accuracy of simulation for the beef sector but also for all other sectors. Improvement of the simulation of nonbeef sectors by partitioning beef is apparently due to the important position of beef in the meat economy. To allow two kinds of beef to enter the wholesale price determination system allows the model to isolate the significantly different direct and cross effect of fed and nonfed beef on each of the other meat sectors. With only one type of beef SIMU V had to estimate a single average relationship and apply this to a heterogeneous beef supply of changing composition.

Using the other criterion, i.e., the average percentage error, the simulation by SIMU VI is slightly less accurate for the cattle-beef sector than that by SIMU V. But, the SIMU VI simulation results for nonbeef sectors and for the overall model are superior to the SIMU V simulations. According to this criterion, the inclusion of fed and nonfed worsened the accuracy in the cattle-beef sector but improved the accuracy for other meat sectors in the system.

Table 8. Percentage Error Indexes and Theil's Inequality Coefficients for SIMU VI and SIMU V Models and for the Sectors Within Each, Calculated From the First Quarter of 1965 Until the Fourth Quarter of 1973a

<table>
<thead>
<tr>
<th></th>
<th>Average percentage error index</th>
<th>Theil's inequality coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SIMU VI</td>
<td>SIMU V</td>
</tr>
<tr>
<td>Beef sector</td>
<td>2.6189</td>
<td>2.5145</td>
</tr>
<tr>
<td>Pork sector</td>
<td>3.3923</td>
<td>4.9069</td>
</tr>
<tr>
<td>Broiler sector</td>
<td>3.8572</td>
<td>4.8187</td>
</tr>
<tr>
<td>Turkey sector</td>
<td>8.0366</td>
<td>8.4848</td>
</tr>
<tr>
<td>Model</td>
<td>3.0184</td>
<td>3.3727</td>
</tr>
</tbody>
</table>

a SIMU VI = Abdou's model.

b SIMU V = model in which beef is treated as a homogeneous product.

Source: Abdou, [11, p. 114].
To any economist, of course, one apparent deficiency of the subsector equations (10) through (14) is the absence of any price variables. Whether the addition of price variables would improve the performance of the subsector, I do not know. But it is a possibility worth looking into.
APPENDIX I: RESTRICTED REGRESSION

Given the model

\[ Y = X\beta + \epsilon \]

where \( X \) is an \( n \times K \) matrix of fixed numbers having rank \( K < n \), \( \epsilon \) is an \( n \times 1 \) vector of errors from a normal independent distribution, \( E\epsilon = 0 \) and \( E\epsilon\epsilon' = \sigma^2 I_n \)

\( Y \) is \( n \times 1 \)

\( \beta \) is \( K \times 1 \)

The classical least-squares (BLUE) estimator of \( \beta \) is

\[ \hat{\beta} = (X'X)^{-1}X'Y \]

The dispersion (variance-covariance) matrix of \( \hat{\beta} \) is

\[ D(\hat{\beta}) = (X'X)^{-1}\sigma^2 \]

Let \( X'_f \) be a (row) vector of independent variables to be used in forecasting \( Y_f \). Then the unbiased forecast is

\[ \hat{Y}_f = X'_f \hat{\beta} \]

and its variance is

\[ D(\hat{Y}_f) = X'_f (X'X)^{-1}X_f \sigma^2 + \sigma^2 = (X'_f D(\hat{\beta})X_f + 1) \sigma^2 \]

Now suppose the vector \( \beta \) is to be estimated subject to the restrictions

\[ R'\beta = r \]

where \( R' \) is a \( J \times K \) known constant matrix of rank \( J < K \) and \( r \) is a \( J \) element known constant vector. The restricted least squares estimator is then

\[ \hat{\beta}_r = \hat{\beta} + F(r - R'\hat{\beta}) = \beta + F(r - R'\beta) + G\epsilon \]

where

\[ F = (X'X)^{-1}R'(R'X'X)^{-1}R \]^{-1} \]

\[ G = (I - FR')(X'X)^{-1}X \]

Assume the constraints are true. Then \( \hat{\beta}_r \) is unbiased and

\[ D(\hat{\beta}_r) = GG'\sigma^2 \]

\[ = D(\hat{\beta}) - D(\hat{\beta})R[R'D(\hat{\beta})R]^{-1}R'D(\hat{\beta}) \]
Because \([R'DR^{-1}]\) is nonnegative definite, diagonal elements of \(D(\hat{\beta}_r)\) are equal to or smaller than the diagonal elements of \(D(\hat{\beta})\). The restricted estimates may have smaller variances, and cannot have larger variances, than the unrestricted estimates. If the constrained estimate is used to forecast \(Y_f\), the unbiased forecast is \(\hat{Y}_f = X'_f\hat{\beta}_r\) and its variance is

\[D(\hat{Y}_f) = D(\hat{\beta}_r) - X'_f FR'(X'X)^{-1}X'_f\sigma^2\]

Because \(FR'(X'X)^{-1}\) is nonnegative definite,

\[D(\hat{Y}_f) \leq D(\hat{\beta}_r)\]

Using constrained estimates can reduce the variance of the forecast, and cannot increase the variance.

Now assume the restrictions are false. Then \(\hat{\beta}_r\) is biased

\[E(\hat{\beta}_r) = \beta + F(r - R'\beta)\]

The mean square error matrix for \(\hat{\beta}_r\) is

\[MSE(\hat{\beta}_r) = F(r - R'\beta)(r - R'\beta)'F + D(\hat{\beta}_r) = (Bias \hat{\beta}_r)(Bias \hat{\beta}_r)' + D(\hat{\beta}_r)\]

Because diagonal elements of \(D(\hat{\beta}_r)\) can be less than diagonal elements of \(D(\hat{\beta})\), mean square errors for biased restricted estimates can be less than variances for unbiased unconstrained estimates. The forecast obtained by using \(\hat{\beta}_r\) is biased

\[E(Y_{fr}) = X'_f[\beta + F(r - R'\beta)]\]

The mean square error for \(Y_{fr}\) is

\[MSE(Y_{fr}) = X'_f[(Bias \hat{\beta}_r)(Bias \hat{\beta}_r)' + D(Y_{fr})]X_f + \sigma^2 = X'_f(MSE \beta_r)X_f + \sigma^2\]

Because \(D(Y_{fr})\) may be less than \(D(\hat{\beta}_r)\), the mean square errors of the biased forecast may be smaller than the variance of the unbiased forecast.

A test of the null hypothesis \(R'\beta = r\) is obtained by using the ratio

\[
\frac{(e'_{r}e_{r} - e'e)/J}{e'e/(n-K)} = F(J, n-K)
\]

which possesses an F distribution with J and n-K degrees of freedom. And
\[ e = Y - X\hat{\beta} \]

\[ e_r = Y - X\hat{\beta}_r \]

Because, as pointed out earlier, restrictions (even incorrect ones) reduce variances of the coefficients, Toro-Vizcarrondo and Wallace [16] propose a test based on a mean square error criterion.

Another possibility is to use stochastic rather than exact restrictions:

\[ R'\hat{\beta} - r = \mu \] where \( \mu \) has zero mean vector and known dispersion matrix. Properties of restricted estimators have been extensively studied by George Judge and his colleagues [3, 9, 10].
APPENDIX II: WEIGHTED REGRESSION

Suppose the \( i \)-th equation in a set of \( G \) equations can be written in vector-matrix notation as \( y_i = Z_i \beta_i + \mu_i \) where \( y_i \) is \( n \times 1 \) and \( Z_i \) is an \( n \times K \) matrix of exogenous variables. The set of all equations can be written as

\[
\begin{pmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_G \\
\end{pmatrix} = \begin{pmatrix}
  Z_1 & 0 & \cdots & 0 \\
  0 & Z_2 & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & Z_G \\
\end{pmatrix} \begin{pmatrix}
  \beta_1 \\
  \beta_2 \\
  \vdots \\
  \beta_G \\
\end{pmatrix} + \begin{pmatrix}
  \mu_1 \\
  \mu_2 \\
  \vdots \\
  \mu_G \\
\end{pmatrix}
\]

\( y = Z\beta + \mu \)

Suppose

\[
E(\mu_j \mu_k') = w_{jk} \sigma^2, \quad w_{jk} \text{ scalar}
\]

Then

\[
E\mu\mu' = W\sigma^2
\]

where \( W \) is an \( nG \times nG \) positive definite matrix. Define the \( nG \times nG \) matrix \( T \) as \( TWT' = I \). Then \( T'T = W^{-1} \). Define, also,

\[
Y = Ty, \quad X = TZ, \quad \varepsilon = Tu.
\]

Premultiplying (A.1) by \( T \) yields

\[
Y = X\hat{\beta} + \hat{\varepsilon}, \quad E\varepsilon\varepsilon' = I\sigma^2
\]

The seemingly unrelated estimate of \( \beta \) obtained by minimizing \( e'e = (Y - Xb)'(Y - Xb) \) is

\[
(A.2) \quad b = (X'X)^{-1}X'Y = (Z'T'TZ)^{-1}Z'T'Ty
\]

If \( Z_1 = Z_2 = \ldots = Z_G = (\text{say}) Z \), then

\[
(A.3) \quad b = \begin{pmatrix}
  \hat{\beta}_1 \\
  \vdots \\
  \hat{\beta}_G \\
\end{pmatrix} = \begin{pmatrix}
  (Z'Z)^{-1}Z'y_1 \\
  \vdots \\
  (Z'Z)^{-1}Z'y_G \\
\end{pmatrix}
\]

If \( w_{jk} = 0 \) for all \( j \neq k \), (3) holds again. In these two circumstances, the seemingly unrelated estimate is simply the set of least squares estimates
obtained by treating each equation separately.

Now define the weighting matrix

\[
P = \begin{pmatrix}
p_1 I & 0 & 0 \\
0 & p_2 I & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & p_g I
\end{pmatrix}, \quad p_i \text{ scalar}
\]

Each I is \(n \times n\); \(P\) is \(nG \times nG\); \(p_i\) is the weight to be assigned to the \(i\)-th dependent variable. A weighted estimate of \(\beta\) is obtained by minimizing \(e' Pe = (Y - Xb_w)' P(Y - Xb_w)\) with respect to \(b_w\). The result is

\[
(A.4) \quad b_w = (Z'T'PTZ)^{-1}Z'T'PTy
\]

Write the transformation matrix \(T\) as \(T = (T_{ij})\ i = 1, 2, \ldots, G;\ j = 1, 2, \ldots, G;\) and \((T_{ij})\) is \(n \times n\); and write \(p_{ij} I = P_{ij}\). Then

\[
Z'T'PTZ = \begin{pmatrix}
Z_1'(\Sigma T_i P_i T_i)Z_1 & Z_1'(\Sigma T_i P_i T_i)Z_2 & \cdots & Z_1'(\Sigma T_i P_i T_i)Z_G \\
Z_2'(\Sigma T_i P_i T_i)Z_1 & Z_2'(\Sigma T_i P_i T_i)Z_2 & \cdots & Z_2'(\Sigma T_i P_i T_i)Z_G \\
\vdots & \vdots & \ddots & \vdots \\
Z_G'(\Sigma T_i P_i T_i)Z_1 & Z_G'(\Sigma T_i P_i T_i)Z_2 & \cdots & Z_G'(\Sigma T_i P_i T_i)Z_G
\end{pmatrix}
\]

\[
Z'T'PTy = \begin{pmatrix}
Z_1'(\Sigma T_i P_i T_i)y_j \\
Z_2'(\Sigma T_i P_i T_i)y_j \\
\vdots \\
Z_G'(\Sigma T_i P_i T_i)y_j
\end{pmatrix}
\]

The weighted estimate of each \(\beta_j\) is a function of all the \(p_i\).

Suppose \(w_{jk} = 0\) for all \(j \neq k\). Then the off-diagonal sub-matrices in \((Z'T'TZ)\) in \((A.2)\) are null matrices because \(Z_j'(\Sigma T_i T_i)Z_k = Z_j'(0)Z_k = 0\). Consequently \((A.3)\) holds. The off-diagonal submatrices in \(Z'T'PTZ\) are \(Z_j'(\Sigma T_i P_i T_i)Z_k\) and these need not be null matrices. Thus, even though errors in different equations are independent, the weighted estimate of each \(\beta_j\) is a function of all the variables in the system and of all the weights.
REFERENCES


