MODEL VALIDATION AND THE NET TRADE MODEL

By William E. Kost*

In this article I discuss validation of structural economic models I emphasize goodness-of-fit measures for historical simulations plus comparisons with alternative models I then use these procedures to evaluate the world trade forecast modeling system being developed in the International Economics Division, ESCS

The article discusses processes of validating structural forecasting models It summarizes methods of evaluating the goodness of fit of model simulations over historical periods and methods of comparing the forecasting behavior of structural models with that of simple time series models The net trade model provides a case study for these two validation processes

Keywords

Validation
Modeling
Forecasting
International trade
Wheat
Coarse grain

VALIDATION

A common approach to analyzing economic issues involves developing a model that simulates economic behavior This model becomes a proxy for reality The model's behavior is then evaluated to provide insight into analyzing economic issues Used this way, models have to represent reality accurately One determines whether or not a model is good through the process known as validation Determining the "goodness" of a model is a subjective process that involves using both economic and statistical criteria One usually begins to construct and validate a model by defining the economic problem that model will analyze This procedure restricts the model's size and scope to only relevant aspects of economic behavior

Once the problem has been identified, an initial structural hypothesis can be proposed General statements are developed concerning the form of the structural equations, the availability of data, structural shifts over time, and the signs and magnitudes of coefficients An appropriate sampling and equation estimation procedure is defined and the preliminary model is estimated These initial equations are evaluated on the basis of both the prior economic hypotheses and statistical, econometric criteria In light of this evaluation, several equations may have to be made more accurate through an alternative equation specification (and/or possibly estimation procedure) that is also consistent with the set of hypotheses previously specified In some instances this equation evaluation leads to rejection of the previously specified hypotheses The prior hypothesis framework must then be redefined and new equations specified and estimated that will be consistent with the new hypotheses This process may also lead to the rejection of the data base, which then requires the generation of a new data base that will lead to different, more acceptable model parameters

An initial model is constructed with this process of hypothesis generation, data base construction, equation estimation, and equation evaluation Only after these steps have been taken can we evaluate the behavior of the complete model How does the complete model track within the historical period of the sample? How does it respond to shocks? How does the model forecast outside the period of the sample?

Simulation Methods

The purpose of model validation is to increase one's confidence in the ability of the model to provide useful information Attention focuses on goodness of fit of the complete model (as opposed to goodness of fit of any single equation) Therefore, model validation continues throughout model construction and even into model use

Tracking the model through the historical period of fit allows evaluation of interdependence between its equations The lowest level of interdependence in any historical simulation is the residual check Under a residual check simulation, all equations are assessed with all explanatory variables set at their actual values For example, assume the model can be represented by a set of n equations

$$Y_t = F(Y_t, Y_{t-1}, \ldots, Y_{t-i}, X_t, X_{t-1}, \ldots, X_{t-j}, \xi_t) \quad (1)$$

*The author is an agricultural economist in the International Economics Division, ESCS

where

- n = the number of endogenous variables in the model,
- m = the number of exogenous variables in the model,
- t = the time period,
- Y = an n column vector of endogenous variables,
- i = the maximum number of lag periods on endogenous variables,
- X = an m column vector of exogenous variables,
- j = the maximum number of lag periods on exogenous variables, and
- ξ = an n column vector of errors

A residual check simulation would be the solution of

$$\hat{Y}_t = \hat{F}(Y_t, Y_{t-1}, \ldots, Y_{t-i}, X_t, X_{t-1}, \ldots, X_{t-j}) \tag{2}$$

for $\hat{Y}_t$ Estimates of model parameters, $\hat{F}$, and actual values for all right-hand variables are used in this calculation This is equivalent to solving each equation independently It provides a check on the accuracy of the solution algorithm The residuals

$$\hat{\xi}_t = Y_t - \hat{Y}_t \tag{3}$$

will be identical to those produced in the econometric estimation of F

The next level of interdependence involves solving

$$\hat{Y}_t = \hat{F}(\hat{Y}_t, Y_{t-1}, \ldots, Y_{t-i}, X_t, X_{t-1}, \ldots, X_{t-j}) \tag{4}$$

In this static simulation, all exogenous and lagged endogenous variables are set at actual values This provides a series of simultaneous solutions for endogeneous variables, each for a single time period Static simulation errors will typically be larger than those in a residual check as this simulation allows for interactions among current-period endogenous variables

A dynamic simulation provides the highest level of interdependence The dynamic simulation involves solving

$$\hat{Y}_t = \hat{F}(\hat{Y}_t, \hat{Y}_{t-1}, \ldots, \hat{Y}_{t-i}, X_t, X_{t-1}, \ldots, X_{t-j}) \tag{5}$$

where only exogenous variables and the initial i period endogenous variables are set at actual values The first time period simulated will have the same solution as the static simulation The second time period will differ, its simulation will use values of the estimated lagged endogenous variables from the previous period, $\hat{Y}_{t-1}$, rather than the actual values, Y_{t-1} The third time period simulated will use estimated endogenous variable values for the first two time periods, and so on throughout the simulation horizon

The dynamic simulation furnishes a simultaneous solution that starts at an initial point in time, based on a set of initial conditions, then feeds on itself for additional inputs throughout the simulation time horizon A dynamic simulation differs from a static simulation, it generates a single multiperiod simulation rather than a series of single-period simulations All multiperiod forecasts of future behavior are dynamic simulations These forecast simulations, of course, also require forecasted, rather than actual, values for the exogenous variables [1] Dynamic simulation errors will typically be larger than those in a static simulation Errors can be propagated throughout the system both by interactions among current-period endogenous variables and by interactions among current and lagged endogenous variables

Each of the three simulations can be evaluated for goodness of fit As a residual check simulation yields information identical to that from the econometric evaluation of individual equations, this article will focus on static and dynamic simulations

Validating Multivariable Models

Problems arise in evaluation of models that simulate many endogenous variables simultaneously Virtually no techniques exist for overall goodness-of-fit evaluation of multiple-response simulation models One can sometimes circumvent this multiple-response problem, either by viewing a simulation with many responses as many simulations, each with a single response, or by combining several responses and treating the combination as a single response

[1] A dynamic simulation that forecasts future behavior involves solving $\hat{Y}_t = \hat{F}(\hat{Y}_t, \hat{Y}_{t-1}, \ldots, \hat{Y}_{t-i}, \hat{X}_t, \hat{X}_{t-1}, \ldots, \hat{X}_{t-j})$

The mathematician says that 2+2 is identically equal to 4 The statistician says that 2+2 is approximately 4 The economist asks, "What kind of number are you looking for?"

Oral tradition

Wallace suggests that "if the question that promotes the research relates to a specific variable, the research should be keyed on that variable Reliability of the model should be based upon how well the key variable is predicted" (*12*, p 15) [2] By their nature, models contain several variables that are relatively unimportant However, the problems for which models are typically used require more than one key variable Wallace's approach narrows the range of focus but still leaves a subjective decision regarding a model's goodness of fit

GOODNESS-OF-FIT MEASURES

Several goodness-of fit measures are now presented for each endogenous variable To the extent they are favorable, they increase one's subjective confidence in the model, and help evaluate changes in the model A comparison of prechange and postchange simulations, in terms of these goodness-of-fit measures, provides information concerning the merit of the structural change Five types of goodness-of-fit measures will be examined errors, regression, correlation, inequality coefficients, and turning points

Errors

Several alternative measures of simulation error can be calculated They all measure the deviation of a simulated variable from the actual path The simplest measure is mean error

$$\text{Mean error} = \frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t) \quad (6)$$

where

T = the number of periods simulated,
$\hat{Y}_t$ = the simulated level of the variable at time period t, and
Y_t = the actual level of the variable at time period t

The mean error can be misleading Large positive and negative errors offset each other and bias the mean error downward

The mean absolute error (MAE) is defined as

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|\hat{Y}_t - Y_t| \quad (7)$$

The mean absolute error is not subject to the bias associated with the mean error

Probably more frequently used in the literature is the root-mean-square (RMS) error

RMS error =

$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2} \quad (8)$$

This measure weights large errors more than the mean absolute error

These three errors can best be evaluated relative to the average size of the variable They, therefore, become more relevant expressed in percentage terms

mean percentage error =

$$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\hat{Y}_t - Y_t}{Y_t}\right) \quad (9)$$

mean absolute relative error (MARE) =

$$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{|\hat{Y}_t - Y_t|}{Y_t}\right) \quad (10)$$

RMS percentage error =

$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\hat{Y}_t - Y_t}{Y_t}\right)^2} \quad (11)$$

In all cases, the smaller the error, the better the fit

Regression

A linear regression of actual values of a variable on predicted values has been suggested by Cohen and Cyert (*1*, pp 112-127) as a method of testing goodness of fit

$$Y_t = \beta_0 + \beta_1 \hat{Y}_t + \xi_t \quad (12)$$

$\hat{Y}_t$ would equal Y_t for all t in perfect models and the resulting regression is one with zero intercept ($\beta_0 = 0$) and unit slope ($\beta_1 = 1$) Parameters of the regression would be tested to see if they differed significantly from zero and one and if the ξ_t's are small

Correlation Coefficient

Association between predicted and actual values for a variable can be measured by the correlation coefficient (R) or by R-square R-square measures the proportion of

[2] Italicized numbers in parentheses refer to items in References at the end of this article

the variation explained by a linear regression of predicted on actual values A disadvantage of the R or R-square as the sole measure of goodness of fit is that perfect correlation only implies an exact linear relationship between predicted and actual values For simulations to be unbiased, and therefore perfect, regression parameters of $\beta_0 = 0$ and $\beta_1 = 1$ must also exist

Theil's Inequality Coefficients

Theil proposed the inequality coefficient as a measure for analyzing accuracy Several definitions of the inequality coefficient exist in the literature Even Theil presents different definitions at different points The first inequality coefficient was proposed by Theil in *Economic Forecasts and Policy* (*11*, pp 32-33) [3]

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\hat{Y}_t^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}Y_t^2}} \quad (13)$$

This inequality coefficient is bounded by zero and one When $U = 0$, $\hat{Y}_t = Y_t$ for all periods, and a perfect simulation exists When $U = 1$, either the model always predicts zero for nonzero actual values, or the model predicts nonzero values for actual values that are always zero, or negative proportionality exists between predicted and actual values Unlike the correlation coefficient, this inequality coefficient penalizes a consistent bias in the simulation However, again unlike the correlation coefficient, it is sensitive to additive transformation of variables [4] When one is evaluating alternative variations of a single model, where general levels of endogenous variables remain relatively unchanged, this disadvantage is not a serious drawback However, this version of the inequality coefficient may not be comparable across models

To overcome sensitivity to an additive transformation, Theil proposed defining the inequality coefficient in terms of changes in a variable The base from which all predicted and actual variables are measured is fixed, and comparisons can then be made across models This inequality coefficient is defined as

$$U_1 = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}((\hat{Y}_t - Y_{t-1}) - (Y_t - Y_{t-1}))^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_{t-1})^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t - Y_{t-1})^2}} \quad (14)$$

[3] This is the definition used in the FEDEASY "Actfit" comparison of actual and predicted time series FEDEASY refers to the set of linkules added to SPEAKEASY by the Federal Reserve System SPEAKEASY is a software package widely used for analysis by ESCS economists

[4] Adding a constant, k, to any set of predicted and actual values will reduce the value of this inequality coefficient by increasing the denominator and leaving the numerator of the fraction defining U unchanged

$$\sqrt{\sum_{t=1}^{T}((\hat{Y}_t + k) - (Y_t + k))^2} \Big/ \sqrt{\sum_{t=1}^{T}(\hat{Y}_t + k)^2} + \sqrt{\sum_{t=1}^{T}(Y_t + k)^2}$$

$$= \sqrt{\sum_{t=1}^{T}(\hat{Y}_t + k - Y_t - k)^2} \Big/ \sqrt{\sum_{t=1}^{T}(\hat{Y}_t + k)^2} + \sqrt{\sum_{t=1}^{T}(Y_t + k)^2}$$

$$= \sqrt{\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2} \Big/ \sqrt{\sum_{t=1}^{T}(\hat{Y}_t + k)^2} + \sqrt{\sum_{t=1}^{T}(Y_t + k)^2}$$

$$< \sqrt{\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2} \Big/ \sqrt{\sum_{t=1}^{T}(\hat{Y}_t)^2} + \sqrt{\sum_{t=1}^{T}(Y_t)^2}$$

The Math-Econ make exquisite modls finely carved from the bones of walrus Specimens made by their best masters are judged unequalled in both workmanship and raw material by a unanimous Econographic opinion If some of these are "useful"—and even Econ testimony is divided on this point—it is clear that this is purely coincidental in the motivation for their manufacture

Axel Leyonhufvud
"Life Among the Econ"

The U_1 inequality coefficient also ranges between zero and one with $U_1 = 0$ occurring when a perfect simulation exists U_1 is always less than U as only the denominator changes from one formulation to the other [5]

Theil proposed a third inequality coefficient in *Applied Economic Forecasting* (*10*, pp 26-29)

$$U_2 = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}((\hat{Y}_t - Y_{t-1}) - (Y_t - Y_{t-1}))^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t - Y_{t-1})^2}} \qquad (15)$$

This U_2 inequality coefficient ranges from zero to infinity For a perfect simulation, when $\hat{Y}_t = \hat{Y}_t$ for all periods, $U_2 = 0$ A no-change forecast model, where $\hat{Y}_t = \hat{Y}_{t-1}$ for all periods, generates a U_2 inequality coefficient of 1 No upper bound on the U_2 inequality coefficient means that there can be a model that is worse than a no-change forecast model

Regardless of the definition chosen for the inequality coefficient, the numerator remains unchanged It is the RMS error defined in equation (8)

The square of the RMS error can be decomposed into several terms, each reflecting a different type of error

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2 = (\hat{Y} - \overline{Y})^2 + (s_{\hat{Y}} - s_Y)^2 + 2(1 - r)\, s_{\hat{Y}} s_Y \qquad (16)$$

where

$$\overline{\hat{Y}} = \frac{1}{T}\sum_{t=1}^{T}\hat{Y}_t$$

$$\overline{Y} = \frac{1}{T}\sum_{t=1}^{T}Y_t$$

$$s_{\hat{Y}} = \sum_{t=1}^{T}(\hat{Y}_t - \overline{\hat{Y}})^2$$

$$s_Y = \sum_{t=1}^{T}(Y_t - \overline{Y})^2$$

$$r = \frac{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - \overline{\hat{Y}})(Y_t - \overline{Y})}{s_{\hat{Y}} s_Y}$$

The first term is zero only when the means of actual and predicted variables are equal Errors that lead to a positive value for this term can be interpreted as a bias or central tendency error The second term is zero only when standard deviations of actual and predicted variables are equal A positive value for this term can be interpreted as error due to different variation The third term is zero only when the correlation coefficient between predicted and actual values is one Therefore, a positive value for this term can be interpreted as an error due to different covariation To compare different model decompositions, one should convert the three components to proportional terms by dividing each by their sum

$$U(\text{bias}) = \frac{(\overline{\hat{Y}} - \overline{Y})^2}{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2} \qquad (17)$$

$$U(\text{variation}) = \frac{(s_{\hat{Y}} - s_Y)^2}{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2} \qquad (18)$$

$$U(\text{covariation}) = \frac{2(1 - r)s_{\hat{Y}}s_Y}{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2} \qquad (19)$$

This implies that

$$U(\text{bias}) + U(\text{variation}) + U(\text{covariation}) = 1 \qquad (20)$$

---

[5]The equivalence of the numerator is demonstrated as follows

$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}((\hat{Y}_t - Y_{t-1}) - (Y_t - Y_{t-1}))^2} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_{t-1} - Y_t + Y_{t-1})^2} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2}$$

One expects U(bias) to be low If it is large, the average errors are large and considerable bias exists in the simulation Even if the inequality coefficient cannot attain its optimum level of zero, the most desired level for U(bias) remains zero One would like U(variation) to be low If U(variation) is high, predicted and actual values have unequal standard deviations This might suggest that the model structure (or equation) underlying the variable in question is misspecified The expectations regarding U(covariation) differ It is unlikely that any model can generate simulations that are perfectly correlated with actual outcomes, therefore, one cannot expect U(covariation) to be low As simulations will not all be perfect, the goal should be the lowest inequality coefficient possible with a decomposition showing U(bias) and U(variation) approaching zero and U(covariation) approaching one With this type of decomposition, systematic error is minimized

An alternative decomposition of the square of the RMS error, that is

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2 = (\bar{\hat{Y}} - \bar{Y})^2 + \qquad (21)$$

$$(s_{\hat{Y}} - rs_Y)^2 + (1 - r^2)s_Y^2$$

can be evaluated in relation to the regression equation defined in equation (12) The first term is the same as that in equation (16) A perfect simulation generates a regression equation with zero intercept and unit slope

$$Y_t = \hat{Y}_t + \xi_t \qquad (22)$$

Because ξ_t has zero mean by definition, Y must equal $\hat{Y}$ and the first term of the decomposition becomes zero Furthermore, the regression slope in equation (12) can be defined as

$$\beta_1 = \frac{\sum_{t=1}^{T}(\hat{Y}_t - \bar{\hat{Y}})(Y_t - \bar{Y})}{\sum_{t=1}^{T}(\hat{Y}_t - \bar{\hat{Y}})^2} = \frac{rs_Y}{s_{\hat{Y}}} \qquad (23)$$

For a perfect simulation, β_1 equals one and this second decomposition term also becomes zero These three components can also be converted to proportional terms

$$U(\text{mean}) = U(\text{bias}) = \frac{(\bar{\hat{Y}} - \bar{Y})^2}{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2} \qquad (24)$$

$$U(\text{regression}) = \frac{(s_Y - rs_Y)^2}{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2} \qquad (25)$$

$$U(\text{residual}) = \frac{(1 - r^2)s_Y^2}{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2} \qquad (26)$$

with

$$U(\text{bias}) + U(\text{regression}) + U(\text{residual}) = 1 \qquad (27)$$

The objective is to generate a model with the lowest inequality coefficient possible for each variable, the decomposition should show U(bias) and U(regression) approaching zero and U(residual) approaching one In fact, if the two decomposition terms differ significantly from zero, a linear correction factor[6] can be applied that will generate the desired decomposition

Turning Point Errors

Another important goodness-of-fit measure is how well actual turning points are simulated during the historical period Turning points are important because many economic time series exhibit positive serial correlation For a model to be superior to a simple time trends model, it must predict turning points

A simulation, with respect to turning points, has four possible outcomes A turning point will actually exist and the model will either predict or not predict it, or no turning point will exist and the model will either predict or not predict one These four possibilities are illustrated in the following diagram

[6] The optimal linear correction factor to $\hat{Y}_t$ will be of the form $\hat{\beta}_0 + \hat{\beta}_1 \hat{Y}_t$ where $\hat{\beta}_1 = rs_Y / s_{\hat{Y}}$ and $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{\hat{Y}}$

ECONOMETRICS ANONYMOUS
One of the major trends of the past decade has been the proliferation of redundant and useless econometric models and analyses This new professional body has been formed to enable an economist, when he feels the urge to run multiple regressions far into the night, to telephone a fellow member of E A who will come over and sit up with him until the desire to regress passes

Leonard Silk
"New Remedies for Economists"

		Predicted	
		Turning Point	No Turning Point
Actual	Turning Point	f_{11}	f_{12}
	No Turning Point	f_{21}	f_{22}

Each cell represents the frequency of each alternative Perfect turning point forecasting implies $f_{12} = f_{21} = 0$, that is, no turning point errors If f_{12} or f_{21} are not equal to zero, turning point errors are occurring Expressing these errors in proportional terms provides a measure of turning point error

A turning point error can be defined as

TP error =

$$\frac{f_{12} + f_{21}}{f_{11} + f_{12} + f_{21} + f_{22}} \quad (28)$$

A measure of error due to turning points missed is

$$TP_M \text{ error} = \frac{f_{12}}{f_{11} + f_{12}} \quad (29)$$

A measure of error due to falsely predicted turning points is

$$TP_F \text{ error} = \frac{f_{21}}{f_{21} + f_{11}} \quad (30)$$

Each of these measures ranges between zero and one, small values indicate good turning point simulations

NET TRADE MODEL VALIDATION

The world trade forecast modeling system under development in the International Economics Division (IED), ESCS, centers on net trade models The net trade model accounts for the interaction among major trading countries by commodity Each commodity model is a system of export supply and import demand functions, by country, that are solved simultaneously for net trade (exports and imports) and world price levels The net trade functions are specified as functions of own price, other commodity prices, production, income, population, and other demand shifters Net trade models for individual commodities are linked through cross price variables I evaluate the wheat and coarse grain net trade models here These models were developed to support the USDA world trade/ U S export outlook process (*6*)

Static Simulation Results

A static simulation of wheat and coarse grain net trade models was performed over 1964-75 The data base for equation specification is, in general, 1960-75 The EEC wheat threshold price series used in the EEC-6 wheat consumption equation starts in 1964 This limitation shortens the simulation period to 1964-75

Table 1 presents a summary of several measures of goodness of fit for each variable Generally, the mean absolute relative errors and root-mean-square percentage errors are low, the slope coefficients of the regressions of actual on predicted values are close to one, the R-square coefficients are high, and the Theil inequality coefficients are low

All U_2 inequality coefficients are substantially below one, therefore, this model proved to be significantly better than a no change forecast model A decomposition of the inequality coefficients shows that errors are due primarily to differences in the covariation between actual and predicted values All U(covariation) and U(residual) terms approach one and other components approach zero Turning points for all variables are, in general, forecasted accurately

In terms of the key variables concept suggested by Wallace, these models were developed to forecast U S agricultural commodity trade within the context of an integrated world agricultural commodity trade model Therefore, U S exports and world price level forecasts become key variables The MARE's for U S wheat (USWHEX) and coarse grain (USCGRX) exports are 4 7 and 4 6 percent, respectively The corresponding RMS percentage errors are 5 9 and 5 5 percent The MARE's for U S wheat (PXWHEUS) and corn (PXCORUSG) prices are 9 4 and 4 2 percent, respectively The corresponding price RMS percentage errors are 12 0 and 5 3 percent

TYPICAL RESULTS ARE SHOWN   *the best results are shown*

CORRECT WITHIN AN ORDER OF MAGNITUDE   *wrong*

Table 1 — Summary of goodness of-fit statistics for each endogenous variable, 1964 75 static simulation

Model and variable[1]	Mean[2]	MARE	RMS percentage error	β_1	R^2	Theil's inequality coefficients			TP error
						U	U_1	U_2	
Wheat net trade model									
ARWHEXC	2787	12 5	14 0	1 12	0 95	0 06	0 17	0 33	0 08
AUWHEXC	7109	6 6	7 4	1 08	.88	04	12	25	08
BRWHEM	2439	5 1	6 5	1 03	93	03	08	16	08
CAWHECON	4522	1 9	2 2	85	87	01	27	57	33
CAWHEEK	15151	9 4	10 5	1 00	93	05	24	50	0
CAWHEX	11991	12 1	13 7	82	51	07	34	71	08
DEWHENM	1076	20 8	24 4	90	84	11	14	26	33
DEWHESTK	2493	13 8	16 6	71	53	08	27	50	17
EGWHENM	2682	3 6	4 1	1 00	.96	02	17	33	08
E6WHECON	29855	2 1	2 7	1 00	75	01	35	62	25
E6WHEEK	5615	18 2	21 6	49	38	10	38	76	25
E6WHEMW	2620	11 7	14 1	1 03	82	07	17	32	08
E6WHEXW	4876	4 5	5 7	1 09	96	03	10	19	08
FRWHENX	5524	11 0	12 8	1 15	91	06	18	35	17
INWHENM	4527	4 1	5 8	1 01	99	03	02	04	0
IRWHENM	483	24 1	28 4	1 01	92	10	05	09	08
JPWHEM	4661	2 4	3 6	90	96	02	27	52	0
KRWHEM	1288	6 8	8 4	1 09	96	04	27	57	0
LAWHENMC	1940	6 3	8 3	97	91	04	28	52	08
NAWHEM	1839	8 3	10 2	1 00	94	04	16	34	25
PKWHENM	1317	11 7	14 6	95	77	07	20	39	17
PXWHEUS	2 34	9 4	12 0	.97	.94	05	24	50	17
RWWHENM	17155	3 7	4 8	1 03	96	02	09	19	17
UKWHEM	4205	6 8	7 8	88	76	04	19	39	08
USWHEX	22776	4 7	5 9	1 04	.95	03	11	21	08
Coarse grain net trade model									
ARCGRNX	5942	8 2	9 4	0 95	0 87	0 05	0 15	0 28	0 17
AUCGRNX	1599	13 3	15 7	1 04	94	07	19	41	17
DECGRNM	4374	7 1	9 0	71	63	04	20	40	08
ESCGRNM	2891	12 2	15 1	1 09	70	07	28	52	25
FRCGRNX	5167	9 6	12 4	1 05	91	06	21	43	0
ITCGRNM	5780	4 4	5 5	84	76	03	18	35	17
JPCGRNM	9773	5 1	6 5	99	95	03	25	47	08
PXCORUSG	1 79	4 2	5 3	1 00	.98	02	13	24	08
RWCGRNM	14265	6 8	8 5	1 00	96	04	24	46	25
SACGRCON	5753	4 4	6 0	82	83	03	53	80	25
SACGRNX	1926	18 2	21 8	1 01	89	09	07	13	0
SVCGRNM	2166	20 5	29 8	99	98	06	07	13	33
THCGRX	1652	5 0	6 0	95	97	03	10	20	17
UKCGRNM	3710	5 6	7 1	93	53	04	34	61	17
USCGRX	26672	4 6	5 5	1 00	98	03	12	23	0

[1] Variables are defined in table 8

[2] Quantities are 1,000 metric tons except for PXWHEUS and PXCORUSG which are in dollars per bushel

THREE OF THE SAMPLES WERE CHOSEN FOR DETAILED STUDY the results on the others didn't make sense and were ignored

INTUITIVELY OBVIOUS I don't understand it either

Oral tradition

Table 2 presents goodness of-fit measures for the overall model These measures are for the integrated wheat and coarse grain model and for both subcomponents of that model separately To calculate the slope coefficient, R-square, Theil inequality coefficients, and turning point relative errors, I assume that the responses of the several variables can be combined and treated as the response of a single variable This approach seems inappropriate for the MARE and RMS percentage error For these measures, a simple average of respective individual endogenous variable measures is reported Thus, the static simulation of the wheat and coarse grain model generally exhibits an 8 8-percent average MARE and a 10 9-percent average RMS percentage error The regression slope coefficient, R-square, and Theil inequality coefficient, including its decomposition, all indicate unbiased forecasts and explain a major proportion of the variation in the actual variables throughout the historical period of fit

Table 2 — Goodness-of-fit statistics wheat and coarse grain net trade models, 1964 75 static simulation

Goodness of fit statistic	Wheat and coarse grain net trade model	Wheat net trade model component	Coarse grain net trade model component
Average MARE	8 8	8 9	8 6
Average RMS percentage error	10 9	10 6	11 4
β_1	1 00	1 00	1 01
R^2	99	99	99
Theil inequality coefficients			
U	03	03	03
U_1	05	05	09
U_2	09	09	17
U(covariation)	9973	9987	9831
U(residual)	9999	9985	9910
Turning point errors			
TP	13	13	14
TP_M	12	10	14
TP_F	14	13	14

Dynamic Simulation Results

A dynamic simulation of the wheat and coarse grain net trade model was performed over 1964-75 Table 3 presents goodness of-fit measures for each variable These results are similar to those from the static simulation Some variables perform slightly worse, and others, slightly better

The U_2 inequality coefficient for South African coarse grain consumption (SACGRCON) exceeds one This signifies that the dynamic forecast for this variable is significantly worse than a simple no-change model forecast This poor performance is easily explained South African consumption in the model is a function of lagged consumption Between 1963 and 1964, actual coarse grain consumption increased 30 percent Therefore, starting the simulation in 1964 creates a large forecast error, one carried through all periods of the simulation The static simulation does not have this large error as the 1965 forecast depends on actual consumption levels in 1964 rather than on levels forecast for 1964 As South African coarse grain net exports (SACGRNX) are a function of consumption, these large consumption errors generate large net export errors These errors would have been substantially reduced had any year other than 1964 been chosen for starting the dynamic simulation

Korean wheat imports (KRWHEM) exhibit a U_2 inequality coefficient close to one This simulation forecasts a more rapid rise in Korean imports throughout the mid sixties than actually occurred Because actual growth was slow in the earlier years of the simulation, a no-change forecast, on the average, would have proved more accurate However, imports did double during the sixties and the model picked up this phenomenon The other goodness-of-fit measures indicate that the Korean wheat import simulation is satisfactory

Table 3 — Goodness-of-fit statistics for each endogenous variable, 1964 75 dynamic simulation

Model and variable[1]	MARE	RMS percentage error	β_1	R^2	Theil's inequality coefficients			TP error
					U	U_1	U_2	
Wheat net trade model								
ARWHEXC	12 5	14 0	1 12	0 95	0 06	0 17	0 33	0 08
AUWHEXC	6 7	7 8	1 20	89	04	12	23	08
BRWHEM	5 4	6 8	1 03	92	03	08	16	0
CAWHECON	1 9	2 3	77	88	01	30	63	08
CAWHEEK	7 7	10 3	1 02	94	05	23	52	0
CAWHEX	13 0	14 4	77	47	07	36	75	17
DEWHENM	20 9	23 7	1 00	84	10	12	23	33
DEWHESTK	12 2	16 5	74	51	08	25	47	08
EGWHENM	3 7	4 2	1 01	96	02	18	34	08
E6WHECON	2 1	2 7	1 00	75	01	35	62	25
E6WHEEK	14 4	19 2	58	41	09	30	59	25
E6WHEMW	12 5	15 3	1 07	79	07	17	33	25
E6WHEXW	12 4	15 4	92	68	07	25	44	17
FRWEENX	9 8	13 3	1 12	90	06	16	30	08
INWHENM	4 3	5 9	1 03	99	03	03	06	0
IRWHENM	29 4	36 1	1 01	87	13	15	27	08
JPWHEM	3 0	4 0	87	96	02	29	57	0
KRWHEM	8 6	11 4	1 11	95	05	35	92	0
LAWHENMC	6 5	8 6	94	91	04	30	57	08
NAWHEM	8 3	10 2	1 00	94	04	16	34	25
PKWHENM	14 2	17 3	89	71	09	28	52	0
PXWHEUS	13 2	15 3	1 06	90	07	30	62	17
RWWHENM	5 8	7 0	1 12	93	03	15	32	17
UKWHEM	7 5	8 4	94	71	04	22	42	17
USWHEX	5 4	6 5	1 10	95	03	13	23	08
Coarse grain net trade model								
ARCGRNX	9 6	10 8	0 98	0 82	0 05	0 17	0 31	0 08
AUCGRNX	12 5	14 9	1 06	95	06	17	35	08
DECGRNM	7 1	9 0	71	63	04	20	40	08
ESCGRNM	12 1	14 8	1 10	72	07	27	51	25
FRCGRNM	11 8	15 7	1 02	86	07	22	42	0
ITCGRNM	4 4	5 5	83	77	03	17	33	25
JPCGRNM	5 1	6 5	99	95	03	25	47	08
PXCORUSG	5 9	7 2	1 02	97	03	18	32	17
RWCGRNM	8 6	11 3	98	94	05	31	62	33
SACGRCON	18 8	20 1	1 81	88	11	78	1 99	75
SACGRNX	86 7	93 2	71	86	29	33	94	25
SVCGRNM	25 9	32 4	93	99	06	11	22	25
THCGRX	5 0	6 0	95	97	03	10	20	17
UKCGRNM	5 7	7 3	91	51	04	35	62	25
USCGRX	7 7	8 9	1 00	97	04	20	38	08

[1] Variables are defined in table 8

The MARE's for U S wheat (USWHEX) and coarse grain (USCGRX) exports are 5 4 and 7 7 percent, respectively The corresponding RMS percentage errors are 6 5 and 8 9 percent The MARE's for U S wheat (PXWHEUS) and corn (PXCORUSG) prices are 13 2 and 5 9 percent, respectively The corresponding price RMS percentage errors are 15 3 and 7 2 percent Relative to the static simulation, these errors are larger, especially for U S coarse grain exports and wheat prices

Table 4 presents several goodness-of fit measures for the complete model Generally, the dynamic simulation of the wheat and coarse grain model exhibits an 11 7-percent average MARE and a 17 5-percent average RMS percentage error The other measures indicate that the dynamic model forecasts are un biased and explain a major proportion of actual variation throughout the historical period of fit [7]

Table 4 — Goodness-of-fit statistics, wheat and coarse grain net trade models, 1964 75 dynamic simulation

Goodness-of-fit statistic	Wheat and coarse grain net trade model	Wheat net trade model component	Coarse grain net trade model component
Average MARE	11 7	9 7	15 1
Average RMS percentage error	17 5	17 4	17 6
β_1	1 01	1 00	1 04
R2	99	99	98
Theil inequality coefficients			
U	04	04	05
U_1	15	15	26
U_2	27	27	62
U(covariation)	9795	9967	8854
U(residual)	9924	9945	9234
Turning point errors			
TP	15	12	21
TP_M	12	09	16
TP_F	16	13	22

Cross Simulation Comparison

The above evidence focuses on goodness of fit of a particular model simulation Table 5 compares the goodness-of-fit measures across the three kinds of simulations discussed Only two measures are broadly comparable across the residual check, static, and dynamic simulation the R-square and the coefficient of variation The R-square for the residual check simulation is derived from the ordinary least squares estimation procedure The R-square for the static and dynamic simulation comes from the regression of actual on predicted values In all three cases, I derive the coefficient of variation by dividing the standard error of the regression by the mean of the dependent variable The residual check simulation measures in table 5 are not completely comparable to the static and dynamic simulations The regression equations are generally fitted over slightly longer time periods The data, however, broadly indicate behavior across simulations The residual check generally performs somewhat better than the static simulation, and the static simulation performs somewhat better than the dynamic simulation These results are expected, each simulation allows for an additional source of errors

VALIDATION THROUGH COMPARISON WITH ALTERNATIVE MODEL SPECIFICATION

Validation questions essentially refer to a model's goodness of fit The validation results discussed earlier are absolute measures of this They all measure the degree of goodness of fit relative to an ideal the perfect forecast model

Another way to validate a model is to compare its specification with those of other models Two types of simple models are good candidates for comparison The first is the no-

[7] The high U_2 inequality coefficient for the coarse grain component is the result of the errors for South Africa explained previously

Table 5 — Three simulations of wheat and coarse grain net trade models

Model and variable[1]	R-square			Coefficient of variation		
	Residual check	Static simulation	Dynamic simulation	Residual check	Static simulation	Dynamic simulation
Wheat net trade model						
ARWHEXC	0 85	0 95	0 95	25	15	15
AUWHEXC	80	88	89	11	8	9
BRWHEM	88	93	92	9	7	7
CAWHECON	84	87	88	3	2	3
CAWHEEK	99	93	94	4	11	11
CAWHEX[2]		51	47		15	16
DEWHENM	90	84	84	23	27	26
DEWHESTK	76	53	51	1	18	18
EGWHENM	97	96	96	5	5	5
E6WHECON	75	75	75	3	3	3
E6WHEEK[2]		38	41		24	21
E6WHEMW	88	82	79	16	15	17
E6WHEXW	96	96	68	10	6	17
FRWHENX	94	91	90	15	14	15
INWHENM	98	99	99	7	6	6
IRWHENM	92	92	87	33	31	40
JPWHEM	98	96	96	4	4	4
KRWHEM	93	96	95	1	9	12
LAWHENMC	95	91	91	9	9	9
NAWHEM	93	94	94	16	11	11
PKWHENM	90	77	71	11	16	19
PXWHEUS[2]		.94	90		13	17
RWWHENM	93	96	93	1	5	8
UKWHEM	92	76	71	5	9	9
USWHEX	92	95	95	9	6	7
Coarse grain net trade model						
ARCGRNX	0 87	0 87	0 82	13	10	12
AUCGRNX	93	94	95	20	17	16
DECGRNM	82	63	63	11	10	10
ESCGRNM	81	70	72	21	17	16
FRCGRNX	94	91	86	17	14	17
ITCGRNM	88	76	77	10	6	6
JPCGRNM	98	95	95	1	7	7
PXCORUSG[2]		.98	97		6	8
RWCGRNM	96	96	94	11	9	12
SACGRCON	91	83	88	0	7	22
SACGRNX	93	89	86	1	24	102
SVCGRNM	98	98	99	19	33	36
THCGRX	97	.97	97	1	7	7
UKCGRNM	66	53	51	8	8	8
USCGRX	97	.98	97	8	6	10

[1] Variables are defined in table 8

[2] No measures are available for the residual check simulation These variables are not econometrically estimated but derived from identity equations and the simultaneous nature of the net trade model

An economist can tell you what will happen under any conditions And his guess is liable to be just as good as anybody else's

Will Rogers

change model, which assumes that next year's forecast will be the same as this year's actual level The second type of model is a simple time trends model, where each endogenous variable is estimated as a function of time only If the net trade model is no better than these relatively simple models, it should be rejected as a useful forecasting tool [8] However, the net trade model proved superior to both alternatives

The Theil U_2 statistic for the net trade model provides a comparison to a no-change forecast model Except for the special case of South African coarse grain consumption, all statistics are below 1 0

The time trends model provides an interesting comparison because time trends are popular with fore casters Equations in a trend model can take numerous forms, but for this analysis, I chose a linear time trend With annual data for 1960-75, far too few observations were available for developing even moderately sophisticated equations When evaluating the linear trend results, in all but a few cases which demonstrated relatively rapid rates of exponential growth, there appeared to be no advantage to using other functional forms

Table 6 presents linear trends model results and table 7 shows summary statistics for the net trade and the linear trends models The no-change model/linear trends model comparison can be evaluated solely based on a U_2 inequality coefficient For 7 of 40 variables (CAWHEEK, INWHENM, KRWHEM, PXWHEUS, PXCORUSG, RWCGRNM, and UKCGRNM), the U_2 statistic exceeds 1 0 for the linear trends model, and the no-change forecast is superior U S wheat and corn prices are two vari ables for which a forecast of no change is better than a linear trend forecast The average U_2 inequality coefficient is 0 84 for wheat and 0 80 for coarse grain Thus, the linear trends model is more accurate than the no change model Table 7 indicates that the static simulation of the net trade model is superior to the linear trends model [9] A more conclusive evaluation involves the net trade model dynamic simulation as the basis for comparison Dynamic simulation results put a model in the worst possible light The net trade model again performs better than the linear trends model (table 7)

Rather naive trend models were used Trend model results could probably be improved if a commodity analyst was careful in choosing the appropriate trend equation for each variable rather than routinely choosing the linear form for all the variables When one examines the time series plots for all the endogenous variables in the net trade model, it seems unlikely that any forecast from a reasonably simple time series model will be better than the net trade model's The net trade model also provides the core upon which a detailed, structural world trade modeling system can be built As it provides equal or better forecasts as well as a structural model framework, the net trade model seems superior to any time series approach

CONCLUSION

The only true test of validity involves using the net trade model in an actual forecasting environment However, how well the model represents one's perception of reality, in both structural and historical tracking senses, provides some preliminary validation Although no definitive conclusions based on statistical theory can be drawn from such an analysis, impressions gathered from it can increase one's confidence in the model

[8] The net trade model structure itself may be rejected as a forecast tool but may still be accepted as a valid modeling construct It can still support a world modeling system framework that, with better and more complete country-sector detail, provides both a better forecast and has explicit structural integrity

[9] Regressing actual values on predicted values generated by a linear trend does not provide a basis for comparison When the ordinary least squares procedure minimizes the sum of squared deviations from the equation, an intercept of zero and a slope of one is assured For the same reason, the decomposition of the inequality coefficient provides little information

Table 6 — Summary of goodness of-fit statistics for each endogenous variable, linear trends model

Model and variable[1]	MARE	RMS percentage error	R^2	Theil's inequality coefficients U_1	Theil's inequality coefficients U_2	TP error	Coefficient of variation
Wheat net trade model							
ARWHEXC	40 0	52 6	0 04	0 46	0 88	0 31	56
AUWHEXC	15 7	19 1	15	40	66	12	20
BRWHEM	16 8	21 8	06	44	71	31	23
CAWHECON	3 0	3 5	76	35	62	31	4
CAWHEEK	28 9	35 6	02	59	1 38	44	38
CAWHEX	17 3	20 0	04	45	83	31	21
DEWHENM	38 5	51 7	09	41	70	25	55
DEWHESTK	14 8	19 6	03	41	67	31	21
EGWHENM	9 2	11 4	84	39	71	19	12
E6WHECON	2 7	3 8	63	54	95	38	4
E6WHEEK	15 2	19 9	02	42	71	25	21
E6WHEMW	21 7	25 1	58	40	70	67	27
E6WHEXW	17 6	22 3	65	42	72	06	24
FRWHENX	17 4	22 9	82	40	71	25	24
INWHENM	40 3	46 3	01	60	1 36	30	49
IRWHENM	63 4	85 1	28	44	82	25	91
JPWHEM	4 4	5 7	94	32	54	12	6
KRWHEM	20 7	23 7	74	55	1 16	19	25
LAWHENMC	12 8	16 0	73	43	81	12	17
NAWHEM	29 4	32 7	55	44	80	19	35
PKWHENM	23 0	26 7	05	40	68	25	29
PXWHEUS	29 3	34 8	36	55	1 24	40	31
RWHENM	14 4	19 3	26	48	94	38	21
UKWHEM	9 6	12 4	20	43	74	25	13
USWHEX	16 3	19 8	37	47	89	19	21
Coarse grain net trade model							
ARCGRNX	18 2	22 6	0 54	0 40	0 63	0 25	24
AUCGRNX	37 9	44 1	55	45	88	25	47
DECGRNM	16 3	19 8	22	36	60	25	21
ESCGRNM	22 6	24 7	71	46	88	25	26
FRCGRNM	23 5	29 1	74	40	68	06	31
ITCGRNM	14 4	17 1	48	49	98	06	18
JPCGRNM	5 4	7 1	98	28	52	06	8
PXCORUSG	21 9	27 2	61	55	1 21	40	29
RWCGRNM	25 8	31 3	49	55	1 25	12	33
SACGRCON	2 9	3 9	96	30	53	31	4
SACGRNX	50 8	54 9	09	42	71	19	59
SVCGRNM	202 7	285 5	60	46	74	19	305
THCGRX	10 9	17 9	83	34	55	12	19
UKCGRNM	8 1	10 3	14	50	1 00	31	11
USCGRX	20 2	25 2	68	48	91	31	27

[1] Variables are defined in table 8

Table 7 — Average goodness-of-fit statistics alternative model comparison

Goodness of-fit statistic	Net trade model						Linear trends model		
	Static simulation			Dynamic simulation					
	Wheat	Coarse grain	Wheat and coarse grain	Wheat	Coarse grain	Wheat and coarse grain	Wheat	Coarse grain	Wheat and coarse grain
MARE	8 9	8 6	8 8	9 7	15 1	11 7	20 9	32 1	25 1
RMS percentage error	10 6	11 4	10 9	17 4	17 6	17 5	26 1	41 4	31 8
U_1	19	20	20	22	25	22	45	43	44
U_2	36	38	38	43	54	47	84	80	83
TP	13	21	13	12	21	15	28	24	27

Table 8 — Definitions of endogeneous variables in net trade model

First Field

AF	Africa
AR	Argentina
AU	Australia
BR	Brazil
CA	Canada
DE	West Germany
EG	Egypt
ES	Spain
E6	European Economic Community-6
FR	France
IN	India
IR	Iran
IT	Italy
JP	Japan
KR	Korea
LA	Other South America
NA	North Africa
PK	Pakistan
RW	Rest of the world
SA	South Africa
SV	U S S R
TH	Thailand
UK	United Kingdom
US	United States

Second field

CGR	Coarse grain
COR	Corn
WHE	Wheat

Third field

CON	Consumption
EK	Ending stocks
M	Imports
MW	Imports excluding intraregional trade
NM	Net imports
NMC	Net imports, local crop year
NX	Net exports
STK	Stocks
X	Exports
XC	Exports, local crop year
XW	Exports, excluding intraregional trade

REFERENCES

(*1*) Cohen, Kalman J , and Richard M Cyert "Computer Models in Dynamic Economics " *Quar J Econ* Vol 75, Feb 1961, pp 112-127

(2) Dhrymes, Phoebus J , E Philip Howrey, Saul H Hymans, Jan Kmenta, Edward E Leamer, Richard E Quandt, James B Ramsey, Harold T Shapiro, and Victor Zarnowitz "Criteria for Evaluation of Econometric Models " *Annals of Econ and Social Measurement*, 1/3 (1972), pp 291-324

(3) Howrey, E Philip, Lawrence R Klein, and Michael D McCarthy "Notes on Testing the Predictive Performance of Econometric Models " *International Econ Rev* Vol 15, No 2, June 1974, pp 366-383

(*4*) Johnson, Keith N , and Lawrence R Klein "Error Analysis of the LINK Model " Project LINK Working Paper Number 7, LINK Central, Econ Res Unit, Univ Pennsylvania, Philadelphia, Oct 1975

(*5*) Kost, William E "Validation Criteria for Models with More than One Endogenous Variable " Unpublished paper, May 1973

(*6*) ________, Martin Schwartz, and Anthony Burris "FDCD World Trade Forecast Modeling System " FDCD Working Paper, Econ , Stat , Coop , Serv , U S Dept Agr , Aug 1979

(7) Labys, Walter C *Dynamic Commodity Models Specification, Estimation, and Simulation* Lexington Books, D C Heath and Company, Lexington, Mass , 1973, pp 215-240

(*8*) Leuthold, Raymond M "On the Use of Theil's Inequality Coefficients " *Am J Agr Econ* , Vol 57, No 2, May 1975, pp 344-346

(9) Pindyck, Robert S , and Daniel L Rubinfeld *Econometric Models and Economic Forecasts* McGraw-Hill, New York, 1976, pp 308-335

(*10*) Theil, Henri *Applied Economic Forecasting* Rand McNally, Chicago, 1966, pp 1-36

(*11*) ________ *Economic Forecasts and Policy* North-Holland Publishing Company, Amsterdam, 1961, pp 1-48

(*12*) Wallace, Thomas D "The General Problem of Spatial Equilibrium A Methodological Issue " *Interregional Competition Research Methods* (ed Richard A King) Agr Policy Inst Series 10 School Agr and Life Sci , N C State Univ , Raleigh, 1963