Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 413

October 1991

COMMENTS ON THE EVALUATION OF POLICY MODELS

Clive W.J. Granger and Melinda Deutsch

ABSTRACT

This paper examines the evaluation of models claimed to be relevant for policy making purposes. A number of tests are proposed to determine the usefulness of such models in the policy making process. These tests are applied to three empirical examples.

Key words and phrases: evaluation, forecasts, forecast encompassing, mean square forecast error, policy models.

COMMENTS ON THE EVALUATION OF POLICY MODELS

Clive W.J. Granger and Melinda Deutsch †

§1. Introduction

Applied economic research produces many empirical models of various parts of the economy. The models are evaluated in a variety of ways, some will report the specification search used to reach the final model, some will employ a battery of specification tests looking at missing variables, parameter consistency or heterogeneity, some will use cross-validation on post sample evaluation techniques, and so forth. Discussion of these procedures and some difficulties that arise can be found in the book of readings, Granger (1990). Many applied papers, as well as some theoretical ones, will end with a section on the "policy implications"of the model. These sections rarely emphasize that strictly the policy implications only follow if the model is correct and is the actual data generating mechanism, which is an unreasonably strong assumption. It is also an assumption that cannot be true if one has two competing policy models. Of course models are built for a variety of purposes and some are fairly easy to evaluate or to compare. For example if two models are built to provide forecasts, they can be run in real time, after the date of the construction, and the forecasts compared using some pre-agreed cost function or criterion. This approach is less easy to apply to a policy model. A complete evaluation would require some organization, such as the Federal Reserve, to use the model to decide on policy and then to see how well the policy thus achieved actually performs. It is unlikely that most models can be evaluated in such a manner, and so less ambitious methods are required.

In this paper we start with a given model that is claimed to have been built for policy purposes. We consider what are the implications of this model being used by a policy maker to try to keep a single variable of interest near to a series of target values. To do this, a policy variable has its value chosen by use of the model. As background variables change and as targets alter, so the policy variable will take a series of values. It can be argued that each different policy value represents a new "policy regime" but we would prefer to keep this phrase to indicate more momentous, less frequent events such as a change in the variable of interest (from unemployment to inflation) or of the policy variable (from interest rates to money supply) or of the policy model being utilized. In these cases the Lucas critique becomes relevant and tests of super-exogeneity can be employed, as in Engle and Hendry(1989). This is an important aspect of policy model evaluation which we will not consider here, but see Hoover and Sheffrin (1990). We will not consider also the argument from the rational expectations literature that policy cannot be successful, although that argument could be used to give an interpretation to some of our results.

It has been suggested that modern policy makers do not consider specific targets and so the type of control mechanism considered here is unrealistic. As a counter-example, it may be noted that the finance ministers of the G-7 countries meet twice a year to give targets for ten or so indicators of their economies, for use in international policy coordination; as discussed by Frankel (1990). Other examples are also available. It should be noted that the G-7 targets are not made public, and so cannot be used by most economists in evaluation exercises.

In what follows we will assume that targets exist but are not known. It is also assumed that policy makers are optimizing rather than satisficing. A good discussion of alternative approaches to policy is given by van Velthoven (1990).

§2. The Control Mechanism

Suppose that the proposed policy model takes the form

$$Y_t = a + bC_t + kX_t + e_t \qquad (2.1)$$

where Y_t is the variable that the decision maker is trying to influence, called the variable of interest, such as unemployment, C_t is the variable that the decision maker has available as a policy variable, such as money supply, X_t is a vector of other variables that influence Y_t, and e_t is the residual, taken to be zero mean white noise. (2.1) is often called the *plant equation* in the control literature. For the moment it will be assumed that the coefficients a, b, k are constant, and in

particular do not change as C_t changes. Let T_t denote a *target series*, which is the desired values of Y_t. Denote $ce_t = Y_t - T_t$, the control error and let $S(ce)$ be the cost function of the decision maker, representing the cost of an error ce.

The objective of the policy maker will be taken to be to manipulate C_{t+1} so that Y_{t+1} is as close as possible to T_{t+1} when using a one-step horizon. More precisely the policy maker will chose C_{t+1} to minimize $E_t[S(ce_{t+1})]$. It should be noted that it is generally not possible to equate Y_{t+1} with T_{t+1} because X_t and e_t are random variables that are not perfectly forecastable or controllable. The timing is also important as C_{t+1} and T_{t+1} are *determined* at time t but in a sense do not become operative until time $t + 1$. Replacing t by $t + 1$ in (2.1) and using a least-squares cost function suggests that the forecast of Y_{t+1} made at time t is

$$f_{t,1}^y = a + bC_{t+1} + kf_{t,1}^x \qquad (2.2)$$

if C_{t+1} is known, as it is potentially for the decision maker. Here $f_{t,1}^x$ is the optimum one-step forecast of X_t using information available as time t *plus* C_{t+1}. Requiring this forecast to be equal to the target, gives

$$C_{t+1} = b^{-1}[T_{t+1} - a - kf_{t,1}^X] \qquad (2.3)$$

and then the forecast error will be,

$$ec_{t+1} = e_{t+1} + k(X_{t+1} - f_{t,1}^X).$$

Thus *if* the decision maker were using (2.1) to determine the policy, the values for the policy variable would be given by (2.3). Given a satisfactory method of forecasting X_{t+1} everything else in (2.3) can be taken as known.

To an outsider things are rather different as C_{t+1} will not be known at time t and neither will be T_{t+1}. The first can be forecast but there is often little direct information about targets.

To an outsider, the best forecast of Y_{t+1} will be

$$g_{t,1}^Y = a + bg_{t,1}^C + kg_{t,1}^X$$

according to the model (2.1), which is being investigated, where $g_{t,1}^X$ is the optimum one step forecast of X_{t+1} using just information available at time t, not including C_{t+1}. The forecast error will now be

$$\epsilon_{t+1} = Y_{t+1} - g_{t,1}^Y$$

3

$$= ec_{t+1} + b(C_{t+1} - g^C_{t,1}) + k(f^X_{t,1} - g^X_{t,1})$$

As these two components will be uncorrelated if (2.1) is correct (as otherwise the size of ec_{t+1} could be reduced), one gets

$$E[\epsilon^2_{t+1}] \geq E[ec^2_{t+1}] \tag{2.4}$$

in general. This suggests the following test for evaluating a policy model. The conditional forecast, using information in C_{t+1}, of Y_{t+1} should on average be superior to the unconditional forecast, using just information available at time t to an outsider. Note that both forecasts are constructed directly from the model that is being evaluated. Of course without this constraint, the conditional forecast should always be better than the unconditional forecast, on average, as it is based on more information. It should be noted that equality holds in (2.4) only if $C_{t+1} = g^C_{t,1}$, that is the control variable C is perfectly forecastable from the information set I_t. In this case the decision maker cannot influence C and so this variable is irrelevant as a control. If the model (2.1) is correct it follows from (2.3) that if C is perfectly forecastable, so will be the target series, T_t and also there is no instantaneous causation from C_t to X_t. In these circumstances, the model (2.1) or any other model would not be relevant for control purposes. Thus non-rejection of the null hypothesis of equality in (2.4) is equivalent to rejection of (2.1) as a control model. As this hypothesis can be rejected in many ways, it does not follow that (2.1) *is* useful for policy if the hypothesis is rejected. This evaluation technique is clearly relevant for a model that claims to be a policy model and is illustrated in the next section. It is also briefly discussed in Chong and Hendry (1986).

Nevertheless, (2.3) can be used to estimate the target series as

$$\hat{T}_{t+1} = bC_{t+1} + a + kf^X_{t,1} \tag{2.5}$$

using the observed value of C_{t+1} and the forecast of X_{t+1}. If some of the estimated target values are unrealistic, such as negative unemployment or very high inflation rates, this would clearly imply that the model is inadequate. A more technical test can be used when Y_t is $I(1)$, so that the change series ΔY_t is stationary, say. It was pointed out in Granger (1988) that if a policy control is somewhat successful, then at the very least Y_t and T_t must be cointegrated. It follows that Y_t and $\hat{T}_t$ will be cointegrated, which implies that $Y_t - bC_t - kX_t - a$ is $I(0)$ or stationary, which can be tested using standard unit root tests. This evaluation technique is discussed in section 4.

To summarize this discussion, there are several evaluation tests that can be applied to a model that is claimed to be relevant for policy selection, assumed to be of form (2.1).

Test 1

From the model two (sets of) forecasts can be constructed. The "unconditional" one-step forecast of Y_{t+1} made at time t, is

$$g_{t,1}^Y = a + bg_{t,1}^C + kg_{t,1}^X$$

where $g_{t,1}^C$ is the optimal one-step forecast of C_{t+1} made at time t using the information set $I_t : X_{t-j}, C_{t-j}, j \geq 1$ and similarly for X. "Conditional" one-step forecasts are now made of the form

$$f_{t,1}^Y = a + bC_{t+1} + kf_{t,1}^X$$

where now f^Y and f^X are the optimal one-step forecasts based on the large information set $J_t : I_t, C_{t+1}$ acting as though C_{t+1} is known at time t, which is correct for the decision made in this framework. Forecast errors will result from the two forecasts

$$\epsilon_{t+1} = Y_{t+1} - g_{t,1}^Y$$

and

$$ec_{t+1} = Y_{t+1} - f_{t,1}^Y$$

and the null hypothesis tested is equality of the mean squared errors of ϵ and ec. Assuming that these errors have zero means, the null hypothesis is easily tested using the procedure discussed in chapter 9 of Granger and Newbold (1987). Define

$$D_t = \epsilon_t - ec_t$$

$$S_t = \epsilon_t + ec_t$$

and the null hypothesis is then equivalent to testing $correlation(D_t, S_t) = 0$, which follows by noting that $cov(D_t, S_t) = E[\epsilon_t^2 - ec_t^2]$. The test employed in the next section is to regress S on a constant and D. If the coefficient of D is significantly different from zero, conclude that SSE of the conditional forecasts is significantly different from the SSE of the unconditional forecast; otherwise conclude they are not. This is the test that is emphasized in the following section. Non-rejection of the null hypothesis is equivalent to a rejection of the usefulness of the model (2.1) for policy purposes. As the forecast horizon used by the policy maker is unknown, the test was repeated for horizons, 1,2,3, and 4 of the observational time unit.

Test 2

If the model (2.1) is correctly specified for the control variable, the conditional forecast should forecast encompass the unconditional forecast, as discussed in Chong

and Hendry (1986). This means that the poorer unconditional forecast contains no information that is helpful in improving the quality of the conditional forecast. This has to be true in theory if the forecasts are not being constructed from a given model. The test uses post-sample data to form the regression

$$Y_{t+1} = a_1 + a_2 f_{t,1} + a_3 g_{t,1} + residual$$

and forecast encompassing occurs if $a_2 = 1, a_1 = a_3 = 0$ and the residual is white noise. Similar regressions can be formed for different horizon forecasts but are not reported below.

Test 3

From (2.5), under the assumption that the model (2.1) is correctly specified, the underlying target series can be estimated. The reasonableness of these estimates can be judged by the evaluator. Clearly, this is not a formal test.

Test 4

If the target series is $I(1)$, then Y_t will be $I(1)$, and this can be tested by standard techniques such as augmented Dickey-Fuller tests. It will then follow that $z_t = Y_t - bC_t - kX_t - a$ is $I(0)$, or stationary, with zero mean. Again z_t can be tested for the null of $I(1)$ and, if this null is rejected and z_t has declining autocorrelations, one may accept its stationarity. Note that this is equivalent to testing for cointegration between Y_t, C_t, and X_t but with given coefficients, as determined by the model. If these coefficients were unconstrained and X_t included lagged Y_t, the test would be of little interest as z_t should then always be $I(0)$.

Test 5

In the above test, it is assumed that the target series T_t is unobserved, which is usually the case. If the targets are available, C_{t+1} could be estimated from (2.3), giving $\hat{C}_{t+1}$, and then a regression run of the form

$$C_{t+1} = \alpha + \beta \hat{C}_{t+1} + error$$

and the null hypothesis that the model is correct is tested for the joint requirements that $\alpha = 0$, $\beta = 1$, error = white noise. This test is not considered in the following section.

Test 6

A further test that can be performed, but which is also not considered in the empirical sections, is to ask if some other policy variable (or its unanticipated component) is missing from the model. Thus, for example, in a model relating money

6

supply to unemployment, unanticipated government expenditure can be considered as a possible missing variable using a Lagrange Multiplier test. If such a potential control variable is missing, the original model cannot be used as a successful policy model.

It should be noted that these tests are not necessarily strong ones and represent only necessary conditions that a policy model should possess. The tests are for a single model and may not be helpful in comparing models. It is also assumed that C_t, or a component of it, is a potentially controllable variable. The tests are not inter-related in a simple way as they concentrate on different types of possible mis-specification of (2.1).

To perform tests 1 and 2 it is necessary to form unconditional forecasts of the control variable C_t. This will usually come from a reduced form regression of C_t on various explanatory variables, including lagged C_t. This regression can be thought of as an approximation to a reaction function for C_t. Clearly, if this equation is badly specified then this could be a reason why the null hypothesis considered in the two tests are rejected. In the empirical model considered in the next section, the control variable is taken to be the unanticipated change in money (denoted DMR), and thus its unconditional forecast will just be zero for all horizons.

There is relatively little discussion in econometrics about how to evaluate a policy model as compared to models designed for forecasting or hypothesis testing. In other disciplines there is consideration given to policy evaluation, but it is usually nontechnical, see for example Nagel (1990).

§3. An Application to a Model of the Unemployment Rate.

The first empirical example examines a model of the unemployment rate proposed by Barro and Rush (1980) which is based on the Natural Rate/Rational Expectations hypothesis. Although this model was not formulated for policy making purposes, it will be used to illustrate some of the concepts presented in this paper. The model of the unemployment rate consists of two equations: the money growth equation and the unemployment rate equation. The money growth equation is given by:

$$DM_t = a_0 + \sum_{j=1}^{6} a_{1j} DM_{t-j} + a_2 FEDV_t + \sum_{j=1}^{3} a_{3j} UN_{t-j} + DMR_t,$$

where M is a quarterly average of $M1$, $DM_t = \log M_t - \log M_{t-1}$, $FEDV_t$ is an estimate of the deviation from the normal value of the real expenditure of the federal

7

government, U is the quarterly average unemployment rate, and $UN_t = \log(\frac{U_t}{1-U_t})$. The residual, DMR_t, is the unanticipated change in the money supply, and is taken to be the control variable. Thus, for this example, the above equation is used only to estimate the control variable since it is unknown to the investigator.

The equation explaining movements in the unemployment rate is given by

$$UN_t = b_0 + \sum_{j=1}^{10} b_{1j}DMR_{t-j} + b_2MIL_t + b_3UN_{t-1} + b_4UN_{t-2} + \text{residual},$$

where MIL =(military personnel / male population aged 15-44).

This equation is the plant equation. Note that the future values of MIL_t are taken as known when forming the conditional and unconditional forecasts.

In addition, the following univariate model of the unemployment rate was used as a benchmark for comparison.

$$UN_t = a_0 + a_1UN_{t-1} + a_2UN_{t-2} + a_3UN_{t-3} + \text{residual}.$$

The above equations were estimated using quarterly data. The estimated money growth equation was

$DM_t = .02 \qquad + \qquad .42DM_{t-1} \quad + \quad .05DM_{t-2} \quad + \quad .12DM_{t-3}$
$\quad\quad (3.05) \qquad\qquad\quad (4.94) \qquad\qquad (.55) \qquad\qquad (1.36)$

$\quad\quad - .15DM_{t-4} + \quad .28\ DM_{t-5} - \quad .03DM_{t-6} + \quad .01FEDV_t$
$\quad\quad (1.71) \qquad\qquad (3.21) \qquad\qquad (.31) \qquad\qquad (1.45)$

$\quad\quad + .01UN_{t-1} + \quad .004UN_{t-2} - \quad .009\ UN_{t-3} + \text{residual}$
$\quad\quad (1.15) \qquad\qquad (.21) \qquad\qquad (.87)$

$\quad\quad T=148\ [1952:1-1988:4] \qquad R^2 = .51 \qquad \text{D.W.} = 2.00 \qquad \hat{\sigma}=.00710$

(the modulus of t-values are shown in parentheses). The estimate of Barro's un-

employment equation was

$$UN_t = -.50 - 2.02\, DMR_t - 1.19\, DMR_{t-1} - 2.46\, DMR_{t-2}$$
$$\quad\ (6.03)\quad (2.87)\qquad\quad (1.64)\qquad\qquad (3.42)$$

$$- 2.51\, DMR_{t-3} - 1.67\, DMR_{t-4} - 1.71\, DMR_{t-5}$$
$$\quad (3.27)\qquad\qquad (2.06)\qquad\qquad (2.03)$$

$$- 2.14\, DMR_{t-6} - 1.16\, DMR_{t-7} - 1.00\, DMR_{t-8}$$
$$\quad (2.53)\qquad\qquad (1.40)\qquad\qquad (1.24)$$

$$-.08\, DMR_{t-9} + .04\, DMR_{t-10} - .76\, MIL_t$$
$$\quad (.01)\qquad\qquad (.05)\qquad\qquad (5.52)$$

$$+1.40\, UN_{t-1} - .64\, UN_{t-2} + residual$$
$$\quad (19.64)\qquad (9.10)$$

$$T = 123[1954:3 - 1985:1]\quad R^2 = .98\quad D.W. = 2.06\quad \hat{\sigma} = .04869$$

The univariate model was estimated to be

$$UN_t = -.12 + 1.66\, UN_{t-1} - .85\, UN_{t-2} + .14\, UN_{t-3} + residual$$
$$\quad (2.45)\ (19.59)\qquad (5.67)\qquad (1.65)$$

$$T = 138[1950:4 - 1985:1]\quad R^2 = .96\quad D.W. = 2.01\quad \hat{\sigma} = .06545$$

In order to test the usefulness of this model for policy making purposes the following analysis was performed. First, the conditional and unconditional forecasts were computed for the model using data from 1985:1 to 1988:4. Recall that if the model is useful for policy purposes, then the conditional forecasts should outperform the unconditional forecasts. In order to determine whether this is the case, these forecasts were be compared in a number of ways. First, for test 1, the root mean sum of squared forecast errors (RMSFE) was computed and is displayed in Table 1. Note that for all forecast horizons, the RMSFE of the unconditional forecasts were smaller than the conditional forecasts.

While this suggests that the conditional forecasts do not outperform the unconditional forecasts, it is desirable to test whether the SSE are significantly different from each other.

Unfortunately, some of the assumptions necessary for the validity of this test appear to be violated in this example. In particular, the one-step ahead forecast errors appear to be autocorrelated and this serial correlation is not completely

9

corrected for by assuming the the residuals from the relevant regression are AR(1). A similar problem seems to hold for the 2, 3, and 4 step ahead forecast errors. In some of the cases, the regression results also indicate that the forecast errors are biased. If, for illustrative purposes, the apparent violation of these two key assumptions is ignored, the conclusion drawn from the test results for all forecast horizons is that the SSE of the conditional and unconditional forecasts are not significantly different.

For test 2, the one-step regression was completely unsatisfactory, with $f_{t,1}$ having a negative coefficient(of -1.12) and the residuals containing strong serial correlation, with Durbin-Watson statistic of .35. The results clearly indicate that conditions required from test 2 for the model to be satisfactory for policy purposes are not found.

Test 4 suggests that an additional test of model adequacy should be employed if the variable that the decision maker is trying to influence, Y_t, is I(1): namely, the estimated target from the model under consideration should be cointegrated with Y_t.[1] As a first step in implementing this test, the augmented Dickey-Fuller (ADF) test with two lags and a constant term (see Granger and Newbold (1989) for a discussion of this test) was used to test for stationarity of the unemployment series. The test statistic was 2.95, thus the results are inconclusive as to whether the unemployment series is I(0) or I(1). However, it shall be assumed, for illustrative purposes, that U is I(1) so that the test for cointegration can be performed. [2] The cointegrating regression of unemployment on the estimated target from Barro's model and a constant was run and the ADF test with two lags was used on the residuals from this regression to determine if they were I(0). The t-statistic from the Dickey-Fuller test with a constant term is 19.80, indicating that unemployment and the estimated target from Barro's model are cointegrated. Thus, the unemployment model does not fail this simple test.

For test 3, as a final test of the usefulness of the model for policy purposes, the estimated target series for the model was computed. It does not appear from Figure 1 that these values are unrealistic, so the model does not fail this simple test.

---

[1] For this example, it should be noted that the estimated target was approximated by the in-sample fitted values from the unemployment model since the number of out-of-sample values was too small to perform the unit root test. In addition, UN and its associated target was transformed to levels since the unemployment models considered in this section estimate the I(0) variable, $UN_t = \log(\frac{U_t}{1-U_t})$ where U_t is the quarterly average unemployment rate.

[2] It should also be noted that the coefficients of the lagged terms in the univariate model for UN add almost to one $(1.66 - .85 + .14 = .95)$.

The above results, taken together, strongly suggest that the conditional forecasts are not superior to the unconditional forecasts and, thus, on this basis, we conclude, that the model is not useful for policy purposes. The error autocorrelation inter alia points to model mis-specification, and the latter may be why the forecast tests are what they are.

§4. An Application to Two Models of the Demand for Borrowed Reserves.

The second example examines two models of the demand for borrowed reserves. These models are relevant to policy since the Federal Reserve targets intermediate reserves, and borrowing from the Discount Window has an obvious impact on the reserves. In addition, studies by Keir (1981) and others suggest that the Federal Reserve uses a version of these models. The first model was proposed by Goldfeld and Kane (1966). Briefly, a bank is assumed to have an exogenous reserve need. It can either borrow from the Federal Reserve or from an alternative source. The bank minimizes its total cost of borrowing by choosing the optimal amount of borrowing from each source. The model proposed by Goldfeld and Kane is

$$R_t^B = a_0 + a_1 K_t + a_2 R_{t-1}^B + a_3 \Delta R_t^{UB} + \epsilon_t$$

where R^B is the level of borrowing from the Discount Window, ΔR^{UB} is the change in unborrowed reserves, $K = i_D - i_S$, and i_D is the discount rate, and i_S is the interest rate of the alternative source.

Dutkowsky (1984) extended Goldfeld and Kane's model by arguing that a switching regression model gave a more accurate representation of the behavior of unborrowed reserves. The switching regression model proposed by Dutkowsky is

$$R_t^B = \begin{cases} a_0^L + a_1^L K_t + a_2^L R_{t-1}^B + a_3^L \Delta R_t^{UB} + \epsilon_t^L & \text{if } K \leq K^* \\ a_0^U + a_1^U \log K_t + a_2^U R_{t-1}^B + a_3^U \Delta R_t^{UB} + \epsilon_t^U & \text{if } K > K^* \end{cases}$$

where K^* is an unobservable switching point that needs to be estimated. The discount rate i_D is the control variable throughout this section.

The above models were estimated using seasonally adjusted monthly data. Due to the difference in the sample period and the seasonal adjustment of the data, an additional lag of R^B was found to be significant and included in the regression results. The Goldfeld and Kane model was estimated to be

$$R_t^B = 97.18 \quad + \quad 91.06 \ K_t \quad + \quad 1.03 \ R_{t-1}^B$$
$$(6.02) \qquad (6.19) \qquad\qquad (22.93)$$

$$-.24 \ R_{t-3}^B - .12 \ \Delta R_t^{UB} + \text{residual}$$
$$(6.71) \qquad (4.11)$$

11

$$T = 198[1959:7-1975:12] \quad R^2 = .95 \quad D.W. = 1.89 \quad \hat{\sigma} = 129.03$$

For the Dutkowsky model, the unobservable switchpoint, K^*, which maximized a likelihood function was found using a grid search and estimated to be 0.15. Dutkowsky's estimated model was: for $K_t > .15$

$$R_t^B = \underset{(5.78)}{225.93} + \underset{(4.86)}{107.35} \log(K_t) + \underset{(21.70)}{1.06} R_{t-1}^B$$

$$\underset{(6.66)}{-.27} R_{t-3}^B - \underset{(4.16)}{.15} \Delta R_t^{UB} + \text{residual}$$

and, for $K_t \leq .15$,

$$R_t^B = \underset{(2.87)}{77.45} + \underset{(1.92)}{64.76} K_t + \underset{(6.74)}{.81} R_{t-1}^B$$

$$\underset{(.83)}{-.07} R_{t-3}^B - \underset{(.09)}{.005} \Delta R_t^{UB} + \text{residual}$$

$$T = 192[1960:1-1975:12] \quad R^2 = .95 \quad D.W. = 1.91 \quad \hat{\sigma} = 130.06$$

The univariate model was

$$R_t^B = \underset{(2.86)}{48.44} + \underset{(18.32)}{1.28} R_{t-1}^B - \underset{(1.36)}{.16} R_{t-2}^B - \underset{(2.95)}{.21} R_{t-3}^B + \text{residual}$$

$$T = 200[1959:5-1975:12] \quad R^2 = .93 \quad D.W. = 2.01 \quad \hat{\sigma} = 164.97$$

For test 1 unconditional forecasts of i_D are required. As i_D is $I(1)$, an AR(3) model for Δi_D was estimated as

$$\Delta i_{D,t} = .54\Delta i_{D,t-1} + .10\Delta i_{D,t-3} + whitenoise$$

Thus the unconditional forecast of $i_{D,t+1}$ is the unconditional forecast of $\Delta i_{D,t+1}$ plus $i_{D,t}$. Multiple step forecasts are easily formed.

The interpretation of the results for this example proved to be less straightforward than for the first example and seemed to depend in part on the forecast horizon considered. Following the analysis above, the RMSFE of the conditional and unconditional forecasts was computed for both models for 1976:1 to 1978:12

and is displayed in Table 2. It can be seen that the conditional RMSFE was less than the unconditional RMSFE for some forecast horizons.

Further investigation was needed to determine whether these differences were statistically different. In particular, the test involving the sum and differences of the forecast errors was employed and the results are displayed in Table 3.

The conclusion of test 1 obtained for the Goldfeld and Kane model is that the conditional and unconditional forecasts are not significantly different for any of the forecast horizons. For the Dutkowsky model, the forecast errors for steps one and four were found to be significantly different suggesting, surprisingly, that the unconditional forecasts were superior to the conditional forecasts for those forecast horizons. The significantly smaller unconditional RMSFE is prima facie evidence of model mis-specification, and the latter may be why the forecast tests are what they are.

For test 2, as a further test of the superiority of the conditional forecasts, the conditional and unconditional forecasts were combined using regression (3.1). For a one-step horizons, the estimated parameters were as follows:

Goldfeld and Kane's Model:

$$R_{t+1}^B = 46.34 \quad - \quad 9.41 \; f_{t,1} \; + \; 10.36 \; g_{t,1} \; + \text{residual}$$
$$\qquad\qquad (.78) \qquad (1.50) \qquad (1.64)$$

$$T = 36[1976:1 - 1978:12] \quad R^2 = .70 \quad D.W. = 2.47 \quad \hat{\sigma} = 228.39$$

which is hardly interpretable and

Dutkowsky's Model:

$$R_{t+1}^B = 84.00 \quad + \quad .12 \; f_{t,1} \; + \; .73 \; g_{t,1} \; + \text{residual}$$
$$\qquad\qquad (1.37) \qquad (.06) \qquad (.09)$$

$$T = 36[1976:1 - 1978:12] \quad R^2 = .66 \quad D.W. = 2.54 \quad \hat{\sigma} = 244.31$$

Both of these application of test 2 do not support the usefulness of the models for policy purposes.

Lastly, for test 3, the estimated target series for each model was computed. It does not appear from Figure 2 and 3 that these values are unrealistic.

Taken together, the above results indicate that neither model is useful for policy purposes since the above analysis suggests that the conditional forecasts do not appear to be superior to the unconditional forecasts.

As a final observation it is interesting to note that most of the tests described above can be used to examine the performance of a model during different regimes. For example, an investigator may believe that Dutkowsky's model is only useful for policy making purposes when $K > K^*$. Thus, the SSE of the conditional and unconditional forecasts for the two regimes may be calculated. The results are displayed in Table 4. Again, there is not clear-cut superiority of the conditional forecasts over the unconditional ones.

§5. An Application to a Model for the Demand for Narrow Money in the United Kingdom.

The final example examines a model of the demand for narrow money in the United Kingdom. The model proposed by Hendry and Ericsson (1991) is

$$\Delta(m - p)_t = a_0 + a_1 \Delta p_t + a_2 \Delta(m - p - y)_{t-1} + a_3 R_t^* + a_4 (m - p - y)_{t-1} + \epsilon_t$$

where $\Delta(m-p)_t$ is the growth rate of real money, Δp_t is the inflation rate, $(m-p-y)_t$ is the inverse of velocity, and R_t^* is the learning-adjusted net interest rate. (See Hendry and Ericsson (1991) for a detailed description of the variables). For the sample period 1964:1-1985:2, the above model was estimated to be

$$\Delta(m - p)_t = .02 \quad - \quad .70 \Delta p_t \quad - \quad .19 \Delta(m - p - y)_{t-1}$$
$$(4.64) \quad\quad (5.27) \quad\quad\quad (2.98)$$

$$-.62 R_t^* - .09 (m - p - y)_{t-1} + \text{residual}$$
$$(8.24) \quad\quad (9.94)$$

$$T = 86[1964 : 1 - 1985 : 2] \quad R^2 = .69 \quad D.W. = 2.15 \quad \hat{\sigma} = .01344$$

The estimated univariate model was

$$\Delta(m - p)_t = .26 \Delta(m - p)_{t-1} + .25 \Delta(m - p)_{t-2} + \text{residual}$$
$$(2.45) \quad\quad\quad (2.29)$$

$$T = 86[1964 : 1 - 1985 : 2] \quad R^2 = .16 \quad D.W. = 1.96 \quad \hat{\sigma} = .02170$$

For test 1 unconditional forecasts of R_t^* are required. A model for ΔR_t^* was estimated as

$$\Delta R_t^* = .01 - .13 R_{t-2}^* + whitenoise$$

The unconditional forecast of R_{t+1}^* is the unconditional forecast of ΔR_{t+1}^* plus R_t^*. Multiple step forecasts are easily formed.

14

Following the analysis in the first example, the RMSFE of the conditional and unconditional forecasts was computed for both models for 1985:3 to 1989:2 and is displayed in Table 5. It can be seen that the conditional RMSFE was less than the unconditional RMSFE for all forecast horizons.

Further investigation was needed to determine whether these differences were statistically different. In particular, the test involving the sum and differences of the forecast errors was employed and the results are displayed in Table 5. The conclusion of test 1 obtained for the Hendry and Ericsson model is that the conditional and unconditional forecasts are not significantly different for steps one, two, and three but are significantly different for step four.

For test 2, as a further test of the superiority of the conditional forecasts, the conditional and unconditional forecasts were combined using regression (3.1). For a one-step horizon, the estimated parameters were as follows:

$$\Delta(m-p)_{t+1} = .002 \quad + \quad 1.15 \ f_{t,1} \ - \ .31 \ g_{t,1} \ + \text{residual}$$
$$(.26) \qquad (2.65) \qquad (.62)$$

$$T = 16[1985:3 - 1989:2] \quad R^2 = .56 \quad D.W. = 2.32 \quad \hat{\sigma} = .01316$$

Lastly, for test 3, the estimated target series for each model was computed. The estimated target series and the observed $\Delta(m-p)$ are within 1 to 3 percentage points of each other, which is not unrealistic.

Taken together, the above results indicate that the usefulness of the above model for policy purposes may depend on the forecast horizon used by the policy maker since the above analysis suggests that the conditional forecasts do not appear to be superior to the unconditional forecasts for steps one, two, three but are superior for step four. However, it should be noted that one interpretation of the failure of the forecast tests to reject is that they lack power due to the small sample size of the forecasts.

§6 Conclusion.

Despite the importance of reliable models used in the policy making process, there has been little consideration given to the evaluation of policy models in the econometric literature. This paper discussed a number of tests that could be used for such a purpose. While these tests provide only necessary properties that a policy model should possess, they do aid the decision maker by excluding some inadequate models.

Appendix

Variable	Description	Source
$M1$	$M1$ (1950:1-1958:12)	Banking and Monetary Statistics:Board of Governors of the Federal Reserve System (1976)
$M1$	$M1$ (1959:1-1988:12)	Citibase Series FM1
U	Unemployment Rate, All Workers, Including Resident Armed Forces	Citibase Series LHURR
F	Total Federal Government Expenditures	Citibase Series GGFEX
$DEFLAT$	Implicit Price Deflator: Federal Government	Citibase Series GDGGF
$MALE$	Male Population Aged 15-44	Sum of Citibase Series PANM4, PANM5, PANM6, PANM7, PANM8, and PANM9
CIV	Civilian Population	Citibase Series POPCIV
POP	Total Population Including Armed Forces Overseas	Citibase Series POP

Table 6. Data from the First Example

The above series were converted into quarterly values when appropriate and the following variables were formed.

List of Variables for Example 1
$DM_t = \log M1_t - \log M1_{t-1}$
$UN_t = \log(\frac{.01U_t}{1-.01U_t})$
$FED_t = \frac{F_t}{DEFLAT_t}$
$FEDV_t = \log FED_t - [\log FED]^*_t$
$[\log FED]^*_t = .05[\log FED]_t + .95[\log FED]^*_{t-1}$
$MIL_t = \frac{POP_t - CIV_t}{MALE_t}$

Table 7.

Variable	Description	Source
R^B	Total Borrowings at Reserve Banks	Citibase Series F6CMB
i_D	Discount Rate, Federal Reserve Bank of New York	Citibase Series FYGD
i_S	Federal Funds Rate	Citibase Series FYFF
R^{TOT}	Total Reserves	Citibase Series FZCMRR

Table 8. Data from the Second Example

The above series were seasonally adjusted and the following variables were formed:

List of Variables for Example 2
$K = i_D - i_S$
$R^{UB}_t = R^{TOT}_t - R^B_t$
$\Delta R^{UB}_t = R^{UB}_t - R^{UB}_{t-1}$

Table 9.

17

References

Barro, R.J., and Rush, M. (1980) Unanticipated Money and Economic Activity. In *Rational Expectations and Economic Policy* (Stanley Fischer, Ed.). University of Chicago Press (for Nat. Bur. Econ. Res.), Chicago.

Chong, Y.Y, and Hendry, D.F. (1986) Econometric Evaluation of Linear Macro-Economic Models, *Review of Economic Studies*, LIII: 671-690. (Also, chapter 17 of Granger, (1990)).

Dutkowsky, D. (1984) The Demand for Borrowed Reserves: A Switching Regression Model, *The Journal of Finance*, 69: No. 2, 407-424.

Engle, R.E., and Hendry, D.F.(1989) Testing Superexogeneity and Invariance. Discussion Paper, Economics Department, University of California, San Diego, Paper No. 89-51, San Diego, California.

Frankel, J.A. (1990) International Nominal Targeting: A Proposal for Overcoming Obstacles to Policy Coordination. Working Paper, Economics Department, University of California, Berkeley.

Goldfeld, S.M., and Kane, E.J. (1966) The Determinants of Member Bank Borrowing: An Econometric Study, *Journal of Finance* 21:499-514.

Granger, C.W.J., and Newbold, P.(1987) *Forecasting Economic Time Series* Academic Press, New York.

Granger, C.W.J. (1988) Causality, Cointegration, and Control, *Journal of Economic Dynamics and Control*, 12:551-559.

Granger, C.W.J.,(Ed.) (1990) *Modelling Economic Time Series: Readings in Econometric Methodology*, Oxford University Press.

Hendry, D. and Ericsson, N. (1990) Modeling the Demand for Narrow Money in the United Kingdom and the United States, *European Economic Review* 35, 4, 833-881.

Hoover, M.D. and Sheffrin, S.M. (1990) Causation, Spending, and Taxes: Sand in the Sandbox or Tax Collector for the Welfare State?. Working Paper, Economics Department, University of California, Davis.

Keir, P. (1981) Impact of Discount Policy Procedures on the Effectiveness of Reserve Targeting. *In New Monetary Control Procedures*, Federal Reserves Staff Study, Board of Governors of the Federal Reserve System, Vol.1,February. .

Nagel, S.S. (1990) *Policy Theory and Policy Evaluation* Greenwood Press New York.

Van Velthoven, B.C.J. (1990) The Applicability of the Traditional Theory of Economic Policy *Journal of Economic Surveys* 4:59-88.

Steps Ahead	D		Barro		Univariate
	Coeff.	$|t|$	Cond	Uncond	
One	-.94	.80	.08	.08	.02
Two	.08	.09	.17	.16	.04
Three	.59	.80	.23	.22	.05
Four	.26	.44	.27	.22	.07

Table 1.

Steps Ahead	Goldfeld/Kane		Dutkowsky		Univariate
	Cond	Uncond	Cond	Uncond	
One	.23	.23	.24	.24	.29
Two	.25	.25	.27	.27	.30
Three	.27	.28	.29	.32	.37
Four	.31	.32	.35	.34	.45

Table 2.(Values shown are divided by 1,000)

Steps Ahead	D (Dutkowsky)		D (Goldfeld/Kane)	
	Coeff.	$|t|$	Coeff.	$|t|$
One	5.98	2.19	-3.35	.28
Two	-.71	.65	-1.57	.53
Three	-.47	.67	-.04	.02
Four	-1.21	2.76	-1.41	1.05

Table 3.

Steps Ahead	SSE for $K_t \leq 0.15$		SSE for $K_t > 0.15$	
	Cond	Uncond	Cond	Uncond
One	.28	.24	1.84	1.86
Two	.07	.07	2.55	2.62
Three	.13	.19	2.95	3.44
Four	.18	.24	4.12	4.03

Table 4.

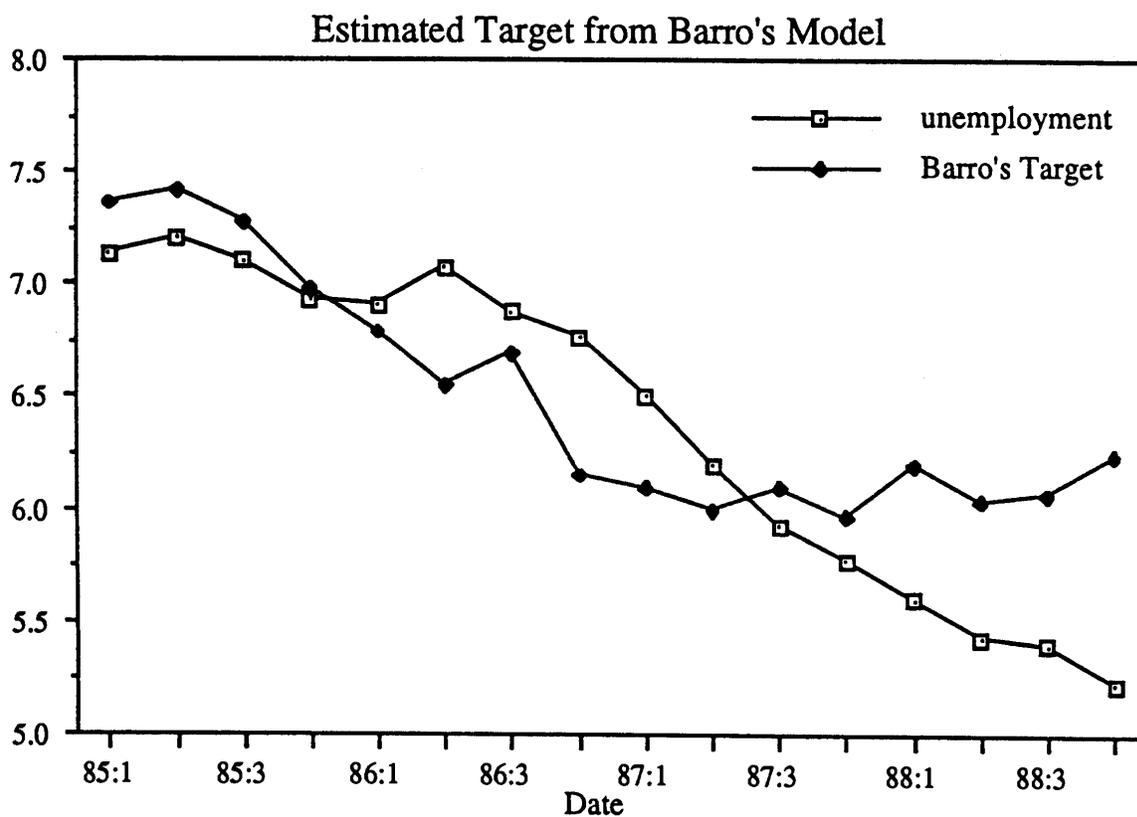| Steps Ahead	D		Hendry/Ericsson		Univariate
	Coeff.	$|t|$	Cond	Uncond	
One	1.34	1.58	1.22	1.52	2.47
Two	.96	1.48	1.66	1.99	2.82
Three	.49	1.09	1.77	2.33	3.21
Four	.83	2.36	1.98	3.20	3.43

Table 5.

Figure 1

Estimated Target from the Goldfeld/Kane Model
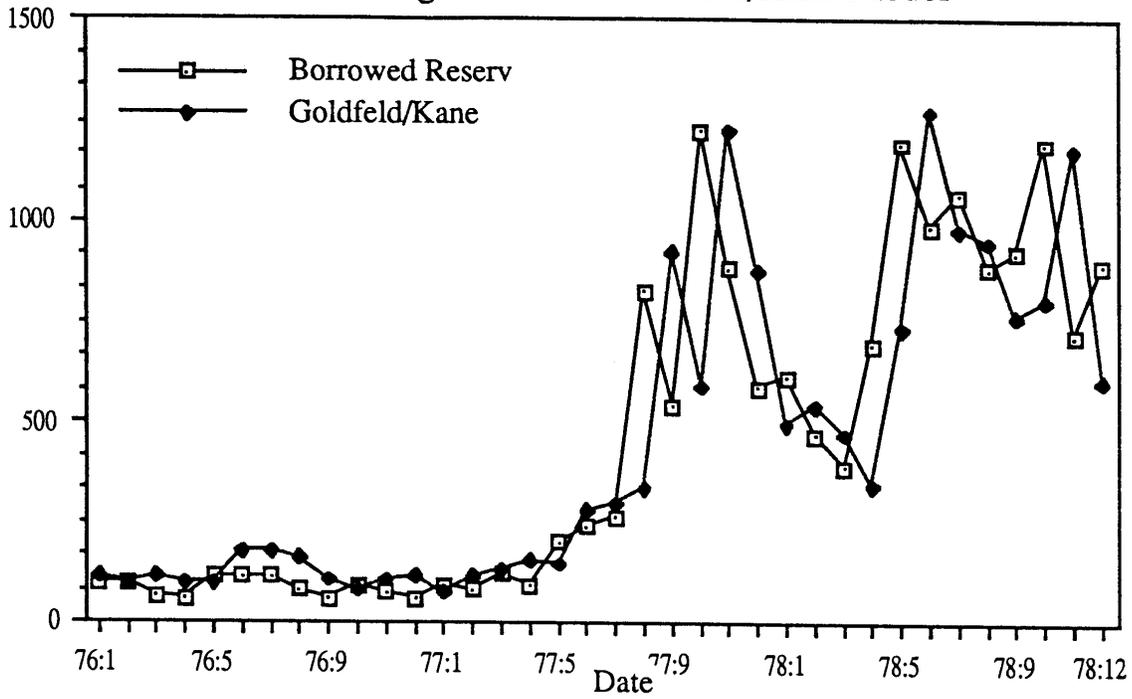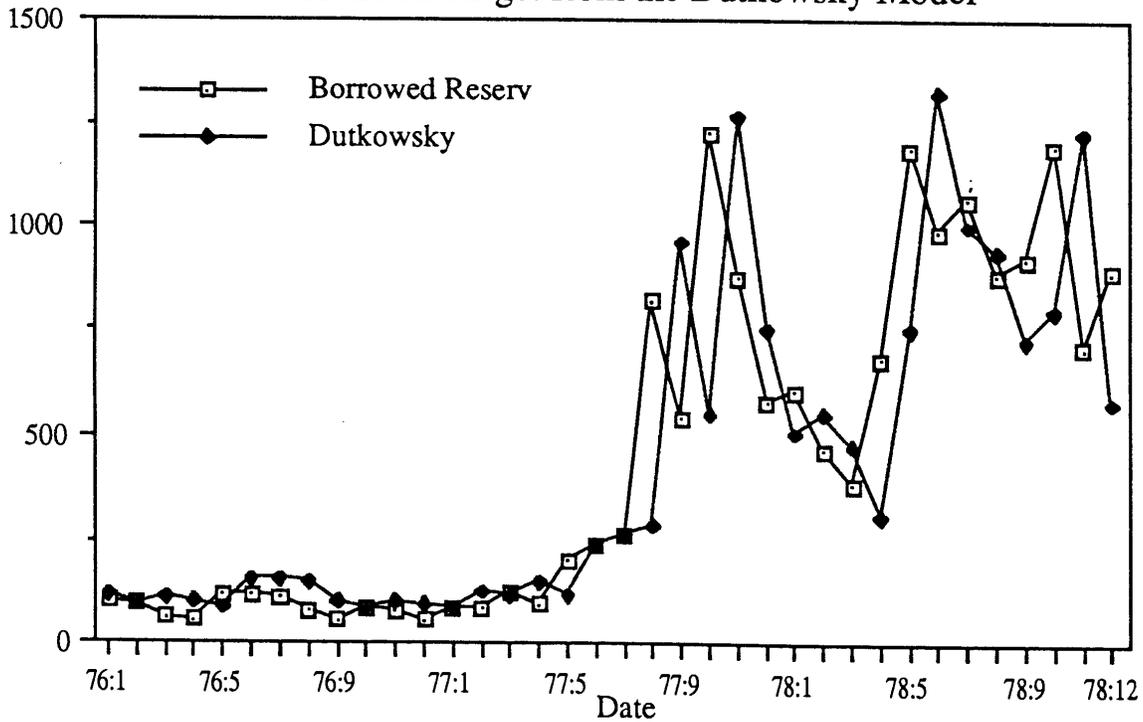

Figure 2

Estimated Target from the Dutkowsky Model

Figure 3

21

IFDP NUMBER	TITLES 1991	AUTHOR(s)
413	Comments on the Evaluation of Policy Models	Clive W.J. Granger Melinda Deutsch
412	Parameter Constancy, Mean Square Forecast Errors, and Measuring Forecast Performance: An Exposition, Extension, and Illustration	Neil R. Ericsson
411	Explaining the Volume of Intraindustry Trade: Are Increasing Returns Necessary?	Donald Davis
410	How Pervasive is the Product Cycle? The Empirical Dynamics of American and Japanese Trade Flows	Joseph E. Gagnon Andrew K. Rose
409	Anticipations of Foreign Exchange Volatility and Bid-Ask Spreads	Shang-Jin Wei
408	A Re-assessment of the Relationship Between Real Exchange Rates and Real Interest Rates: 1974 - 1990	Hali J. Edison B. Dianne Pauls
407	Argentina's Experience with Parallel Exchange Markets: 1981-1990	Steven B. Kamin
406	PC-GIVE and David Hendry's Econometric Methodology	Neil R. Ericsson Julia Campos Hong-Anh Tran
405	EMS Interest Rate Differentials and Fiscal Policy: A Model with an Empirical Application to Italy	R. Sean Craig
404	The Statistical Discrepancy in the U.S. International Transactions Accounts: Sources and Suggested Remedies	Lois E. Stekler
403	In Search of the Liquidity Effect	Eric M. Leeper David B. Gordon
402	Exchange Rate Rules in Support of Disinflation Programs in Developing Countries	Steven B. Kamin
401	The Adequacy of U.S. Direct Investment Data	Lois E. Stekler Guy V.G. Stevens
400	Determining Foreign Exchange Risk and Bank Capital Requirements	Michael P. Leahy
399	Precautionary Money Balances with Aggregate Uncertainty	Wilbur John Coleman II

International Finance Discussion Papers

AUTHOR(s)

1991

398	Using External Sustainability to Forecast the Dollar	Ellen E. Meade Charles P. Thomas
397	Terms of Trade, The Trade Balance, and Stability: The Role of Savings Behavior	Michael Gavin
396	The Econometrics of Elasticities or the Elasticity of Econometrics: An Empirical Analysis of the Behavior of U.S. Imports	Jaime Marquez
395	Expected and Predicted Realignments: The FF/DM Exchange Rate during the EMS	Andrew K. Rose Lars E. O. Svensson
394	Market Segmentation and 1992: Toward a Theory of Trade in Financial Services	John D. Montgomery

1990

393	*Post* Econometric Policy Evaluation A Critique	Beth Ingram Eric M. Leeper
392	Mercantilism as Strategic Trade Policy: The Anglo-Dutch Rivalry for the East India Trade	Douglas A. Irwin
391	Free Trade at Risk? An Historical Perspective	Douglas A. Irwin
390	Why Has Trade Grown Faster Than Income?	Andrew K. Rose
389	Pricing to Market in International Trade: Evidence from Panel Data on Automobiles and Total Merchandise	Joseph E. Gagnon Michael M. Knetter
388	Is the EMS the Perfect Fix? An Empirical Exploration of Exchange Rate Target Zones	Robert P. Flood Andrew K. Rose Donald J. Mathieson
387	Estimating Pass-through: Structure and Stability	William R. Melick
386	International Capital Mobility: Evidence from Long-Term Currency Swaps	Helen Popper
385	Is National Treatment Still Viable? U.S. Policy in Theory and Practice	Sydney J. Key
384	Three-Factor General Equilibrium Models: A Dual, Geometric Approach	Douglas A. Irwin
383	Modeling the Demand for Narrow Money in the United Kingdom and the United States	David F. Hendry Neil R. Ericsson