COMMENTS ON THE EVALUATION OF POLICY MODELS

Clive W.J. Granger and Melinda Deutsch

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors.
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.
§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.

† The first author is a Professor of Economics at the University of California, San Diego and the second author is a graduate student at the same institution. This paper was revised while the first author was a Visiting Scholar at the Federal Reserve Board. This paper represents the views of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or members of its staff. Partially supported by NSF grant SES 89-02950.
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

$$Y_t = a + bC_t + kX_t + e_t$$

(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

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

(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

$$C_{t+1} = b^{-1}[T_{t+1} - a - kf_{t,1}^X]$$

(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

$$\epsilon_{t+1} = Y_{t+1} - g_{t,1}^Y$$
\[ = e c_{t+1} + b(C_{t+1} - g_{t,1}^C) + k(f_{t,1}^X - g_{t,1}^X) \]

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

\[ E[\varepsilon^2_{t+1}] \geq E[ec^2_{t+1}] \quad (2.4) \]

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 \( C_{t+1} = g_{t,1}^C \), 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

\[ \hat{T}_{t+1} = bC_{t+1} + a + kf_{t,1}^X \quad (2.5) \]

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 \( \Delta Y_t \) 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 $\epsilon$ 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 $\text{correlation}(D_t, S_t) = 0$, which follows by noting that $\text{cov}(D_t, S_t) = E[\epsilon_t^2 - ec_t^2]$. 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

\[ Y_{t+1} = a_1 + a_2 f_{t,1} + a_3 g_{t,1} + \text{residual} \]

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

\[ C_{t+1} = \alpha + \beta \hat{C}_{t+1} + \text{error} \]

and the null hypothesis that the model is correct is tested for the joint requirements that \( \alpha = 0, \beta = 1, \) \text{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
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
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

$$UN_t = b_0 + \sum_{j=1}^{10} b_j DMR_{t-j} + b_2 MIL_t + b_3 UN_{t-1} + b_4 UN_{t-2} + \text{residual},$$

where $MIL = \text{(military personnel / male population aged 15-44)}$.

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.

$$UN_t = a_0 + a_1 UN_{t-1} + a_2 UN_{t-2} + a_3 UN_{t-3} + \text{residual}.$$  

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

$$DM_t = .02 + .42 DM_{t-1} + .05 DM_{t-2} + .12 DM_{t-3}$$

$$- .15 DM_{t-4} + .28 DM_{t-5} - .03 DM_{t-6} + .01 FEDV_t$$

$$+ .01 UN_{t-1} + .004 UN_{t-2} - .009 UN_{t-3} + \text{residual}$$

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

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

\[ UN_t = -0.50 - 2.02 DMR_t - 1.19 DMR_{t-1} - 2.46 DMR_{t-2} \]

(6.03) (2.87) (1.64) (3.42)

- 2.51 DMR_{t-3} - 1.67 DMR_{t-4} - 1.71 DMR_{t-5}

(3.27) (2.06) (2.03)

- 2.14 DMR_{t-6} - 1.16 DMR_{t-7} - 1.00 DMR_{t-8}

(2.53) (1.40) (1.24)

- 0.08 DMR_{t-9} + 0.04 DMR_{t-10} - 0.76 MIL_t

(0.1) (0.05) (5.52)

+ 1.40 UN_{t-1} - 0.64 UN_{t-2} + \text{residual}

(19.64) (9.10)

\[ T = 123[1954 : 3 - 1985 : 1] \quad R^2 = 0.98 \quad D.W. = 2.06 \quad \hat{\sigma} = 0.04869 \]

The univariate model was estimated to be

\[ UN_t = -0.12 + 1.66 UN_{t-1} - 0.85 UN_{t-2} + 0.14 UN_{t-3} + \text{residual} \]

(2.45) (19.59) (5.67) (1.65)

\[ T = 138[1950 : 4 - 1985 : 1] \quad R^2 = 0.96 \quad D.W. = 2.01 \quad \hat{\sigma} = 0.06545 \]

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 \( Y_t \). 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. 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, \( UN_t = \log \left( \frac{U_t}{1 + U_t} \right) \) 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 \( R_t^B \) is the level of borrowing from the Discount Window, \( \Delta R_t^{UB} \) 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 \( R_t^B \) was found to be significant and included in the regression results. The Goldfeld and Kane model was estimated to be

\[ R_t^B = 97.18 + 91.06 K_t + 1.03 R_{t-1}^B - .24 R_{t-3}^B - .12 \Delta R_t^{UB} + \text{residual} \]

\[ (6.02) \quad (6.19) \quad (22.93) \]

\[ (6.71) \quad (4.11) \]
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 $K_t > .15$

\[
R_t^B = 225.93 + 107.35 \log(K_t) + 1.06 R_{t-1}^B
\]

\[
(5.78) \quad (4.86) \quad (21.70)
\]

\[-.27 R_{t-3}^B - .15 \Delta R_t^{UB} + \text{residual}
\]

\[
(6.66) \quad (4.16)
\]

and, for $K_t \leq .15$,

\[
R_t^B = 77.45 + 64.76 K_t + .81 R_{t-1}^B
\]

\[
(2.87) \quad (1.92) \quad (6.74)
\]

\[-.07 R_{t-3}^B - .005 \Delta R_t^{UB} + \text{residual}
\]

\[
(.83) \quad (.09)
\]

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

The univariate model was

\[
R_t^B = 48.44 + 1.28 R_{t-1}^B - .16 R_{t-2}^B - .21 R_{t-3}^B + \text{residual}
\]

\[
(2.86) \quad (18.32) \quad (1.36) \quad (2.95)
\]

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

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

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

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:

\[ R^B_{t+1} = 46.34 - 9.41 f_{t,1} + 10.36 g_{t,1} + \text{residual} \]

\[ (0.78) \quad (1.50) \quad (1.64) \]

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

which is hardly interpretable and

Dutkowsky’s Model:

\[ R^B_{t+1} = 84.00 + .12 f_{t,1} + .73 g_{t,1} + \text{residual} \]

\[ (1.37) \quad (0.06) \quad (0.09) \]

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

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, $\Delta p_t$ 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

\[
\begin{align*}
\Delta (m - p)_t &= .02 - .70 \Delta p_t - .19 \Delta (m - p - y)_{t-1} \\
&= (4.64) \quad (5.27) \quad (2.98)
\end{align*}
\]

\[
\begin{align*}
- .62 R^*_t - .09 (m - p - y)_{t-1} + \text{residual} \\
&= (8.24) \quad (9.94)
\end{align*}
\]

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

The estimated univariate model was

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

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

For test 1 unconditional forecasts of $R^*_t$ are required. A model for $\Delta R^*_t$ was estimated as

$$\Delta R^*_t = .01 - .13 R^*_{t-2} + \text{whitenoise}$$

The unconditional forecast of $R^*_{t+1}$ is the unconditional forecast of $\Delta R^*_{t+1}$ 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:

\[
\Delta(m - p)_{t+1} = 0.02 + 1.15 f_{t,1} - 0.31 g_{t,1} + \text{residual}
\]

\[
T = 16[1985 : 3 - 1989 : 2] \quad R^2 = 0.56 \quad D.W. = 2.32 \quad \hat{\sigma} = 0.01316
\]

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U$</td>
<td>Unemployment Rate, All Workers, Including Resident Armed Forces</td>
<td>Citibase Series LHURR</td>
</tr>
<tr>
<td>$F$</td>
<td>Total Federal Government Expenditures</td>
<td>Citibase Series GGFEX</td>
</tr>
<tr>
<td>$DEFLAT$</td>
<td>Implicit Price Deflator: Federal Government</td>
<td>Citibase Series GDGGF</td>
</tr>
<tr>
<td>$MALE$</td>
<td>Male Population Aged 15-44</td>
<td>Sum of Citibase Series PANM4, PANM5, PANM6, PANM7, PANM8, and PANM9</td>
</tr>
<tr>
<td>$CIV$</td>
<td>Civilian Population</td>
<td>Citibase Series POPCIV</td>
</tr>
<tr>
<td>$POP$</td>
<td>Total Population Including Armed Forces Overseas</td>
<td>Citibase Series POP</td>
</tr>
</tbody>
</table>

Table 6. Data from the First Example
The above series were converted into quarterly values when appropriate and the following variables were formed.

<table>
<thead>
<tr>
<th>List of Variables for Example 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$DM_t = \log M1_t - \log M1_{t-1}$</td>
</tr>
<tr>
<td>$UN_t = \log(\frac{.01U_t}{1-.01U_t})$</td>
</tr>
<tr>
<td>$FED_t = \frac{E}{DEFLAT_t}$</td>
</tr>
<tr>
<td>$FEDV_t = \log FED_t - [\log FED]_t^*$</td>
</tr>
<tr>
<td>$[\log FED]_t^* = .05[\log FED]<em>t + .95[\log FED]</em>{t-1}^*$</td>
</tr>
<tr>
<td>$MIL_t = \frac{POP_{t-CIV}}{MALE_t}$</td>
</tr>
</tbody>
</table>

Table 7.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^B$</td>
<td>Total Borrowings at Reserve Banks</td>
<td>Citibase Series F6CMB</td>
</tr>
<tr>
<td>$i_D$</td>
<td>Discount Rate, Federal Reserve Bank of New York</td>
<td>Citibase Series FYGD</td>
</tr>
<tr>
<td>$i_S$</td>
<td>Federal Funds Rate</td>
<td>Citibase Series FYFF</td>
</tr>
<tr>
<td>$R^{TOT}$</td>
<td>Total Reserves</td>
<td>Citibase Series FZCMRR</td>
</tr>
</tbody>
</table>

Table 8. Data from the Second Example

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

<table>
<thead>
<tr>
<th>List of Variables for Example 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K = i_D - i_S$</td>
</tr>
<tr>
<td>$R^{UB}_t = R^{TOT}_t - R^B_t$</td>
</tr>
<tr>
<td>$\Delta R^{UB}_t = R^{UB}<em>t - R^{UB}</em>{t-1}$</td>
</tr>
</tbody>
</table>

Table 9.
References


<table>
<thead>
<tr>
<th>Steps Ahead</th>
<th></th>
<th>( D )</th>
<th></th>
<th>Barro</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>(</td>
<td>t</td>
<td>)</td>
<td>Cond</td>
</tr>
<tr>
<td>One</td>
<td>-.94</td>
<td>.80</td>
<td>.08</td>
<td>.08</td>
<td>.02</td>
</tr>
<tr>
<td>Two</td>
<td>.08</td>
<td>.09</td>
<td>.17</td>
<td>.16</td>
<td>.04</td>
</tr>
<tr>
<td>Three</td>
<td>.59</td>
<td>.80</td>
<td>.23</td>
<td>.22</td>
<td>.05</td>
</tr>
<tr>
<td>Four</td>
<td>.26</td>
<td>.44</td>
<td>.27</td>
<td>.22</td>
<td>.07</td>
</tr>
</tbody>
</table>

Table 1.

<table>
<thead>
<tr>
<th>Steps Ahead</th>
<th>Goldfeld/Kane</th>
<th></th>
<th>Dutkowsky</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cond</td>
<td>Uncond</td>
<td>Cond</td>
<td>Uncond</td>
<td>Univariate</td>
</tr>
<tr>
<td>One</td>
<td>.23</td>
<td>.23</td>
<td>.24</td>
<td>.24</td>
<td>.29</td>
</tr>
<tr>
<td>Two</td>
<td>.25</td>
<td>.25</td>
<td>.27</td>
<td>.27</td>
<td>.30</td>
</tr>
<tr>
<td>Three</td>
<td>.27</td>
<td>.28</td>
<td>.29</td>
<td>.32</td>
<td>.37</td>
</tr>
<tr>
<td>Four</td>
<td>.31</td>
<td>.32</td>
<td>.35</td>
<td>.34*</td>
<td>.45</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Steps Ahead</th>
<th>( D ) (Dutkowsky)</th>
<th></th>
<th>( D ) (Goldfeld/Kane)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>(</td>
<td>t</td>
<td>)</td>
</tr>
<tr>
<td>One</td>
<td>5.98</td>
<td>2.19</td>
<td>-3.35</td>
<td>.28</td>
</tr>
<tr>
<td>Two</td>
<td>-.71</td>
<td>.65</td>
<td>-1.57</td>
<td>.53</td>
</tr>
<tr>
<td>Three</td>
<td>-.47</td>
<td>.67</td>
<td>-.04</td>
<td>.02</td>
</tr>
<tr>
<td>Four</td>
<td>-1.21</td>
<td>2.76</td>
<td>-1.41</td>
<td>1.05</td>
</tr>
</tbody>
</table>

Table 3.

<table>
<thead>
<tr>
<th>Steps Ahead</th>
<th>SSE for ( K_t \leq 0.15 )</th>
<th></th>
<th>SSE for ( K_t &gt; 0.15 )</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cond</td>
<td>Uncond</td>
<td>Cond</td>
<td>Uncond</td>
</tr>
<tr>
<td>One</td>
<td>.28</td>
<td>.24</td>
<td>1.84</td>
<td>1.86</td>
</tr>
<tr>
<td>Two</td>
<td>.07</td>
<td>.07</td>
<td>2.55</td>
<td>2.62</td>
</tr>
<tr>
<td>Three</td>
<td>.13</td>
<td>.19</td>
<td>2.95</td>
<td>3.44</td>
</tr>
<tr>
<td>Four</td>
<td>.18</td>
<td>.24</td>
<td>4.12</td>
<td>4.03</td>
</tr>
</tbody>
</table>

Table 4.
Table 5.

<table>
<thead>
<tr>
<th>Steps Ahead</th>
<th>$D$</th>
<th>Hendry/Ericsson</th>
<th>Univariate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>$</td>
<td>t</td>
</tr>
<tr>
<td>One</td>
<td>1.34</td>
<td>1.58</td>
<td>1.22</td>
</tr>
<tr>
<td>Two</td>
<td>.96</td>
<td>1.48</td>
<td>1.66</td>
</tr>
<tr>
<td>Three</td>
<td>.49</td>
<td>1.09</td>
<td>1.77</td>
</tr>
<tr>
<td>Four</td>
<td>.83</td>
<td>2.36</td>
<td>1.98</td>
</tr>
</tbody>
</table>

Estimated Target from Barro's Model

Figure 1
Estimated Target from the Goldfeld/Kane Model

- Borrowed Reserv
- Goldfeld/Kane

Date

Figure 2

Estimated Target from the Dutkowsky Model

- Borrowed Reserv
- Dutkowsky

Date

Figure 3

21
<table>
<thead>
<tr>
<th>IFDP NUMBER</th>
<th>TITLES 1991</th>
<th>AUTHOR(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>413</td>
<td>Comments on the Evaluation of Policy Models</td>
<td>Clive W.J. Granger, Melinda Deutsch</td>
</tr>
<tr>
<td>412</td>
<td>Parameter Constancy, Mean Square Forecast Errors, and Measuring Forecast Performance: An Exposition, Extension, and Illustration</td>
<td>Neil R. Ericsson</td>
</tr>
<tr>
<td>411</td>
<td>Explaining the Volume of Intraindustry Trade: Are Increasing Returns Necessary?</td>
<td>Donald Davis</td>
</tr>
<tr>
<td>410</td>
<td>How Pervasive is the Product Cycle? The Empirical Dynamics of American and Japanese Trade Flows</td>
<td>Joseph E. Gagnon, Andrew K. Rose</td>
</tr>
<tr>
<td>409</td>
<td>Anticipations of Foreign Exchange Volatility and Bid-Ask Spreads</td>
<td>Shang-Jin Wei</td>
</tr>
<tr>
<td>406</td>
<td>PC-GIVE and David Hendry's Econometric Methodology</td>
<td>Neil R. Ericsson, Julia Campos, Hong-Anh Tran</td>
</tr>
<tr>
<td>405</td>
<td>EMS Interest Rate Differentials and Fiscal Policy: A Model with an Empirical Application to Italy</td>
<td>R. Sean Craig</td>
</tr>
<tr>
<td>404</td>
<td>The Statistical Discrepancy in the U.S. International Transactions Accounts: Sources and Suggested Remedies</td>
<td>Lois E. Stekler</td>
</tr>
<tr>
<td>403</td>
<td>In Search of the Liquidity Effect</td>
<td>Eric M. Leeper, David B. Gordon</td>
</tr>
<tr>
<td>402</td>
<td>Exchange Rate Rules in Support of Disinflation Programs in Developing Countries</td>
<td>Steven B. Kamin</td>
</tr>
<tr>
<td>401</td>
<td>The Adequacy of U.S. Direct Investment Data</td>
<td>Lois E. Stekler, Guy V.G. Stevens</td>
</tr>
<tr>
<td>400</td>
<td>Determining Foreign Exchange Risk and Bank Capital Requirements</td>
<td>Michael P. Leahy</td>
</tr>
<tr>
<td>399</td>
<td>Precautionary Money Balances with Aggregate Uncertainty</td>
<td>Wilbur John Coleman II</td>
</tr>
</tbody>
</table>

Please address requests for copies to International Finance Discussion Papers, Division of International Finance, Stop 24, Board of Governors of the Federal Reserve System, Washington, D.C. 20551.
<table>
<thead>
<tr>
<th>IFDP NUMBER</th>
<th>TITLES</th>
<th>AUTHOR(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>398</td>
<td>Using External Sustainability to Forecast the Dollar</td>
<td>Ellen E. Meade</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Charles P. Thomas</td>
</tr>
<tr>
<td>397</td>
<td>Terms of Trade, The Trade Balance, and Stability: The Role of Savings Behavior</td>
<td>Michael Gavin</td>
</tr>
<tr>
<td>396</td>
<td>The Econometrics of Elasticities or the Elasticity of Econometrics: An Empirical Analysis of the Behavior of U.S. Imports</td>
<td>Jaime Marquez</td>
</tr>
<tr>
<td>395</td>
<td>Expected and Predicted Realignments: The FF/DM Exchange Rate during the EMS</td>
<td>Andrew K. Rose</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lars E. O. Svensson</td>
</tr>
<tr>
<td>394</td>
<td>Market Segmentation and 1992: Toward a Theory of Trade in Financial Services</td>
<td>John D. Montgomery</td>
</tr>
<tr>
<td>393</td>
<td>Post Econometric Policy Evaluation</td>
<td>Beth Ingram</td>
</tr>
<tr>
<td></td>
<td>A Critique</td>
<td>Eric M. Leeper</td>
</tr>
<tr>
<td>392</td>
<td>Mercantilism as Strategic Trade Policy: The Anglo-Dutch Rivalry for the East India Trade</td>
<td>Douglas A. Irwin</td>
</tr>
<tr>
<td>391</td>
<td>Free Trade at Risk? An Historical Perspective</td>
<td>Douglas A. Irwin</td>
</tr>
<tr>
<td>390</td>
<td>Why Has Trade Grown Faster Than Income?</td>
<td>Andrew K. Rose</td>
</tr>
<tr>
<td>389</td>
<td>Pricing to Market in International Trade: Evidence from Panel Data on Automobiles and Total Merchandise</td>
<td>Joseph E. Gagnon</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Michael M. Knetter</td>
</tr>
<tr>
<td>388</td>
<td>Is the EMS the Perfect Fix? An Empirical Exploration of Exchange Rate Target Zones</td>
<td>Robert P. Flood</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Andrew K. Rose</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Donald J. Mathieson</td>
</tr>
<tr>
<td>386</td>
<td>International Capital Mobility: Evidence from Long-Term Currency Swaps</td>
<td>Helen Popper</td>
</tr>
<tr>
<td>385</td>
<td>Is National Treatment Still Viable? U.S. Policy in Theory and Practice</td>
<td>Sydney J. Key</td>
</tr>
<tr>
<td>383</td>
<td>Modeling the Demand for Narrow Money in the United Kingdom and the United States</td>
<td>David F. Hendry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Neil R. Ericsson</td>
</tr>
</tbody>
</table>