# THE USE OF DUMMY IN ECONOMETRIC FORECASTINC

by
ANASTASSIOS GAGALES
Economic Research Department, Bank of Greece

1. Out- of- sample forecasting has become an important ingredient of diagnostic checking. Despite its increasing popularity, however, forecasting exercises are still conducted in a rather laborious and time - consuming fashion which become a deterrent especially when when a large number of alternative specifications is involved.

In the present note we address precisely this problem and we suggest a cunning procedure which (i) enriches the informational content of the standard computer output, (ii) reduces forecasting costs, (iii) introduces a computationally efficient way ot summarizing and testing the forecasting ability of a statistical model and (iv) facilitates the conduct of predictive structural stability tests.

2. Consider the linear model $Y = X\beta + u$, where Y and u denote $\eta$ — dimensional column vectors, $\beta$ is a p — dimensional column vector and X is a np matrix of rank p, with $\rho$ ( $\eta$. Define the kth *observation - specific dummy variable* by the $\eta$ — dimensional coordinate vector $d^k = (di^k)$, where $dk^k$ — 1 and $di^k = 0$ for i different than k. This type of dummy variable has two interesting properties that will come very handy in the sequel :

(i) The inclusion of an observation - specific dummy in a regression is equivalent to the exclusion of the corresponding observation from the samble[1], and

1. Observation specific dummy variables have been used extensively in the presence of outlying observations. They provide, however, a rather extreme solution to the outlier problem. In most instances, the contribution of these non - recurring events can be expressed as c(i) = z(i) f + v (i), where c(i), z(i) and v(i) are coordinate vectors with all but the $i^{th}$ element equal to

(ii) The coefficient of the $k^{th}$ observation - specific dummy is equal to the difference between the actual and the projected value of the dependent variable, in case the sample on which the projection is based does not include the $k^{th}$ observation.

These two properties hold for the general Aitken estimator and for any other estimator isomorphic to it, as in instrumental variable and two - stage least squares estimators, Conchrane - Orcutt and Hildreth - Lu estimation procedures, etc. For expository purposes, and without any loss of generality, the proofwill becasted in terms of the ordinary least squares estimator.

3. The least squares estimator ot $\beta$, when the $k^{th}$ observation is excluded from the sample, is obtained as the solution of the following minimization problem.[2]

$$\min_{[b]} \Sigma_k (Y_i - X_i b)^2 \tag{1}$$

where Xi is a p-dimensional row vector and Yi is a scalar. If, on the other hand, the $k^{th}$ observation is retained in the sample, while at the same time the set of regressors is augmented to include $d^k$, the corresponding least squares estimator can be obtained as the solution of:

$$\min_{[b,f]} \Sigma (Y_i - X_i b - d_i^k f_k)^2 \tag{2}$$

For the purposes of the proof it suffices to show that both (1) and (2) admit

zero ; z(i) denotes the magnitude of the non - recurring event, the scalar f stands for the impact effect while v(i) denotes the measurement error. Finally, c(i) denotes the constant adjustment. It can be shown that as long as E[v(i)] = 0 the presence of measurement errors does not give rise to inconsistent estimates. In a similar fashion one can introduce prior beliefs-information about measurement errors for any one of the explanatory variables. This way of modeling non - sample information corresponds to mixed estimation and can be given a Bayesian interpretation.

2. The following notation is used: Min $[_x]$ F(x) stands fo7 the minimization of F(x) with respect to $\chi$ ; $\Sigma$ denotes the summation over all i, while in $\Sigma_k$ the summation is over all values of i different from k. Finally, $f_k$ and f are used interchangeably.

the same minimun and that this minimum is attained for the same values of b and $f_k$.

Notice first that $min[b,f]$ $(Y_k - X_k b - f)2$ = min [f] $(Y_k - X_k b - f)2$, i.e. for any choice of b one can always chose f so that the squared deviation from the $k^{th}$ observation is zero. This ,in conjunction with the superadditivity of the min(.) functional, implies that :

$$\min[b,f] \ \Sigma(Y\iota - Xi^{i} - di^{k}f)2>$$

$$\min_{[b]\Sigma k}(Yi-Xib)2 + \min_{[b!}f](Yk-Xk^{b}-f)2 = \min_{[b]} \ \Sigma_k(Yi-Xib)2 \quad (3)$$

The RHS of the above inequality denotes the minimum sum of squared residuals when b and f are restricted in the subspace C = (b,f) I Yk = Xkb+f).The LHS on the other hand denotes the corresponding unrestricted minimum. By definition, the unrestricted minimum cannot exceed the restricted one :

$$\min_{[b,f]} \ \Sigma(Y\acute{\iota} \text{ -Xi } dkf)2 <$$

$$\text{minc}[b,\varepsilon] \ \Sigma_k(Yi - X_i b - dt^{x}f)^2 = \min_{[b]} \Sigma_k(Yi - Xib)2 \qquad (4)$$

Since inequalities (3) and (4) hold simultaneously, (3) must hold with equality. This concludes the first part of the proof.

Let b,* f* be the values of b and f that minimize the LHS of (3). Note that the optimal values are related by $f^* = Y_k - X_k b^*$. Let also that RHS of (4) is minimized for some b**. Because of the strict convexity of the minimands, the two optimal solutions [b*, f *'] and b** must be unique. It remains to be shown that b* = b** This can be established easily by contradiction : Let that the LHS of (3) is minimized for some b* different than b**. Then, either $(Yk-X_k b^{**})^2 > 0$, in which case b** cannot be optimal, or the optimal solution is not unique. O.E.D.

In a nutshell, this proposition says that one can obtain the same estimate of β under either one of the following two procedures : (i) least squares estimation with the $k^{th}$ observation excluded from the sample, or (ii) least squares estimation using the entire sample and a dummy for the $k^{th}$ observation. In the latter case, the coefficient of $a^k$ is equal to $f_k = Y_k - X_k b$, namely the forecast error of Yk when b has been estimated from a sample which does not include the $k^{th}$observation. Positive values of $f_k$ imply that the estimated model underpredicts the corresponding observation of the dependent variable.

If the $k^{th}$ observation of the dependent variable is set equal to zero while at the same time $d^k$ is included among the regressors, then minus the coefficient of the observation-specific dummy is equal to the conditional mean of the dependent variable, i.e. = $f_k$ - Xkb. In forecasting and in static out - of- sample simulation experiments one purports to estimate Yk conditional on some senario(s) about the exogenous and the predetermined variables, Xk. These senario(s) can be used to augment the original sample (X,Y) by stacking underneath it the sui - generis observation (Xk,0). Running ordinary least squares with the appropriate observation-specific dummies included among the regressors, one can then obtain, in a single computer run, both the OLS estimate of β and the forecast of the dependent variable, - fk — Xkb.

4. The properties of $f_k$ just menitioned suggest the following four - part computational scheme :

STEP I Decompose the samble (X,Y) in two subsections :
section A, to be used for parameter estimation only ; and
section B, to be used exclusively for testing out - of - sample forecasting performance.

STEP II Append to the original sample (X,Y) a third section, section C, consisting of estimates of the exogenous and the predetermined variables, Xc, for the forecasting horizon.

The corresponding values of the variable to be forecasted are set equal to zero.

STEP III To each observation in sections B and C assign an pobservation-specific dummy, DB and DC respectively.

STEP IV Estimate β using the entire samble, i,e. sections A,B and C together wtth the augmented set of regressors.

$$\begin{pmatrix} Y_A \\ Y_B \\ 0 \end{pmatrix} = \begin{pmatrix} X_A \\ X_B \\ X_C \end{pmatrix} b + \begin{pmatrix} 0 & 0 \\ DB & 0 \\ 0 & DC \end{pmatrix} \begin{pmatrix} f_B \\ f_C \end{pmatrix}$$

5. The above scheme can be used to readily calculate from the standard computer output the Z(q) statistic of forecasting accuracy based on the q post-samble observations[3]. Let σ stand for the standard error of the regression, fi denote the forecasting error in period i, and Si and ti denote respectively the standard error and the t * statistic of fi. Then, Z(q) is defined[4] as :

$$Z(q) = \Sigma(f_i)\ \sigma)^2 = \Sigma t_i^2\ (S_i\ \sigma^2$$

It is distributed asmptotically as chi * square with q degrees of freedom and takes on high values either when ti is high or when Si is large in comparison to S. A significant value of Z(q) indicates either that the model has been incorrectly specified and/or that the stochastic properties of the data generating process have changed.

The null hypothesis Ho : fB = 0 provides an alternative predictive test for structural stability. The test statistic is

$$F = [SSR_B - SSR)\ SSR[\ (n - p)\ /n_B)$$

where SSRB and SSR denote the resticted, $f_B = 0$, and the unrestricted sum of squared residuals, both calculated on the basis of Sections A and B the sample[5]. It is distributed as $F(n_B, \eta - \rho)$, ρ being the number of estimated parameters while $n_B$ denotes the number of observations in Section B. This is a variant of the Chow structural stability test[6] and comes particularly handy when the model cannot be reestimated in Section B because of an insufficient number of observations, i.e. $n_B < p$.

3. See Davidson et al (1978) and Hendry (1980). Another widely used measure of out-of sample forecasting performance is the Mean Squared Forecasting Error, MSFE. It is trivial to show that Z(q) and MSFE are related linearly by Z(q) = q(MSFE S2).

4. The summation runs from T + 1 to T +q.

5. Because ot the presence of observation specific dummies for easch observation in Section B, the unrestricted SSR is identical to the one obtained from Section A alone.

6. Chow (1960), Dufour (1980) and (1982), Gujarati (1970), Harvey (1976), Oaxaga (1974) **and** Valentine (1971).

Both the F and the Z(q) tests evaluate structural stability indirectly on the basis of the model' s predictive performance. They rely on the idea that a correctly specitied and carefully estimated model should not produce consistently off-track forecasts. They do not identidfy sensitive or time - dependent parameters. In this connection it must be noted that specification tests are most useful when (i) the values of the explanatory variables are replicated in the test sample (Section B) or at least lie in the same region, and(ii) the ceteris paribus clause is approximately valid both before and after the shift period. When the explanatory variables move into new regions one cannot separate specification shifts from structural shifts or from the effect of omitted variables.

6. The working as well as the usefulness of the above scheme are perhaps best illustrated in the following example[7] in which we use the 1966- 82 annual samble to (i) estimate a money demand function for Greece, (ii) test for the one-period * ahead forecasting ability of the fitted equation[8], and (iiii) forecast the 1983 level of money demand deposits.

The estimation period is 1986 - 81 (Section A), 1982 has been reserved for static out-of- sample simulation (Section B), while 1983 is the forecast year (Section C). The 1983 value of M4 has been set equal to zero and the sample has been augmented to include 1983 forecasts for the explanatory variables (Step 11). Finally, two dummies, D82 and D83, have been introduced for each one of the two years 1982 and 1983 (Step 111).

Ordinary least squares estimates are presented in the Table overleaf. Z,C, Y, R, EINF, RAT and UHOUS denote respectively the annual average of M4 deflated by CPI, the constant term, GPD at 1970 prices, the interest rate on twelve month time deposits, an unbiased inflation forecast, a ratchet variable measuring the drachma yield of saving accounts in the US, and a measure of credit rationing. All variables are measured in logarithms and hance the residuals are expressed percentage deviations from the corresponding fitted values.

The coefficient of D82 indicates that the 1982 forecast underpredicts the ac-

7. The examble has been chosen solely for exqository purposes. See Gagales (1985) for a detailed study of this specification.

8. In this case static and dynamic forecasts coincide.

tual values by 5,3 %; the coefficient of D83 produces a 1983 forecast for real money balances equal to exp (6.05384) = 425,7 billion Drs. with a standard deviation of 3,4 %. The Z[9](1) statistic is equal to (5.34/1.66)2=10.35, far above its critical valus at the 1% level of significance. On the other hand, the F * statistic[10] is marginally significant at [(0,35009-0,22089)/9,22089] [(17 - 9)/1)]= 4.679. The extention for k>1 is straightforward.

9. To forecast nominal money holding one should either add In (CPI 1983) = 1.6448 to D83 to obtain a forecast of 2205.4 bn drs. or, equivalently, set the 1983 value of Z equal to-ln(CPI 1983) =-1.6448.

10. In view of the identity $F(1,d) = t_d^2$, for d = 1 the above test is equivalent to at - test.

```
ORDINARY LEAST SQUARES

DEPENDENT VARIABLE:   Z
SUM OF SQUARED RESIDUALS =            .220892E-02
STANDARD ERROR OF THE REGRESSION =          .166167E-01
MEAN OF DEPENDENT VARIABLE =          5.21014
STANDARD DEVIATION =              1.36612
R-SQUARED =  0.9999
ADJUSTED R-SQUARED =  0.9999
F-STATISTIC(   9.,     8.) =          12766.3
LOG OF LIKELIHOOD FUNCTION =          55.5097
NUMBER OF OBSERVATIONS =        18.
SUM OF RESIDUALS =           .104904E-04
DURBIN-WATSON STATISTIC (ADJ. FOR  0. GAPS) =  2.5218
```

| RIGHT-HAND VARIABLE | ESTIMATED COEFFICIENT | STANDARD ERROR | T-STATISTIC |
|---|---|---|---|
| C | -4.06337 | .973314 | -4.17478 |
| Z1 | .396098 | .177726 | 2.22870 |
| Y | .541000 | .331253 | 1.63319 |
| Y1 | .805860 | .361713 | 2.22790 |
| EINF | -1.71350 | .218759 | -7.83281 |
| R | .440110 | .314128 | 1.40105 |
| RAT | -.515254E-01 | .241306E-01 | -2.13527 |
| UHOUS | -.336521E-01 | .385863E-01 | -.872125 |
| D82 | .534015E-01 (a) | .246871E-01 | 2.16313 |
| D83 | -6.05384 (b) | .335846E-01 (c) | -180.257 |

(a) 1982 forecasting error (percentage).
(b) 1983 forecast (logarithm of)
(c) Sampling error of the 1983 forecast (percentage)

1 Sample Output

## REFERENCES


1. Brown, R.L. J. Durbin and J.M. Evans (1975), «Techniques for Testing the Constancy of Regression Relationships Over Time", JRSS, Series B, vol 37, pp. 149 -192.
2. Chow, G.C. (1960), «Tests of Equality Between Sets of Cofficients in Two Linear Regressions", Econornetrica, Vol 28, pp. 591-603.
3. Dufour, J.M. (1980), «Dummy Variables and Predictive Tests for Structural Change», Economics Letters, Vol 6, pp. 241 - 247.
4. Dufour, J.M. (1982), «Generalized Chow Tests for Structural Change : A Coordinate Free Approach»,I1ER, Vol 23, pp. 565 - 575.
5. Gagales, A. (1985), «The Demand for Money in Greece», Ph.D. dissertation,University of Rochester.
6. Gujarati, D. (1970), «Use of Dummy Variables in Testing for Equality Between Sets of Coefficients in Linear Regression : A Generalization», The American Statistician, Vol 24, No 4, pp. 18-22.
7. Harvey, A.C. (1974), «An Alternative Proof and Generalization of a Test for Structural Change», The American Statistician, Vol 30, pp. 122- 123.
8. Oaxaga, R.I. (1974), «Another look at Tests of Equality Between Sets of Coefficients in Two Linear Regressions», The American Economist, Vol 18, pp. 23-32.
9. Velentine, T.J. (1971), «The Chow Test : A Comment», The American Statistician, Vol 25, No 1, p.p 62 - 63.