Scand. J. of Economics 93(2), 213–217, 1991

Comment on G. E. Mizon, "Modelling Relative Price Variability and Aggregate Inflation in the United Kingdom"

Herman K. van Dijk

Erasmus University, Rotterdam, The Netherlands

In most cases simple correlations between economic variables do not provide enough evidence to confirm an economic theory and justify a particular policy. Nowadays, this statement hardly needs any defense. Stylized facts have to be interpreted in a serious manner. Superficial analysis may lead to absurd conclusions. A nice anecdote is quoted in Fisher's (1966, pp. 2–3) book on identification. In Tsarist Russia there was a positive correlation between the geographical distribution of the occurrence of cholera and the geographical distribution of the presence of doctors, since the government had sent doctors to the affected areas. The story goes that distraught farmers in the affected areas killed the doctors. The danger of superficial theory and superficial causal inference remains, even when modern methods are used. Suppose we begin with the superficially simple assumption that doctors' behavior is such that they rush to areas where there are sick people. If we summarize this behavior in a mathematical formula, make use of simulation techniques to generate artificial data from this model and compare them with the real data, we obtain a naive confirmation that doctors are present where diseases occur and that doctors should be prevented from going to areas where people are sick. Clearly, in order to determine the cause of a disease, we need a more profound theory that includes other relevant variables. Moreover, a simple predictive analysis over a relatively short period would have refuted the naive theory.

In order to overcome the problems of superficially simple inference, Mizon advocates the *general to specific* modeling strategy where extensive use is made of *instrumental variable* estimation and several different *testing* procedures. These methods are collected in the computer package PC-GIVE (General Instrumental Variable Estimation). My comments refer to these topics.

Mizon's suggestion that the general to specific modeling strategy is good, while the specific to general modeling strategy is seriously flawed, requires some interpretation. The advantage of the general to specific modeling approach is that, given an initial information set with respect to some economic variables, the learning process may be organized in a systematic way and some robustness may be achieved against errors of specification. In the analysis of macroeconomic time series, an attractive sequence of steps in the learning procedure is to study first the time series properties (univariate and multivariate), next the exogeneity properties and thereafter the overidentification properties; see e.g. Monfort and Rabemananjara (1990) and Hendry and Mizon (1989). However, in microeconometrics there is usually a more structured set of initial hypotheses on the agents' behavior. Similarly, in markets of homogeneous goods such as agricultural and financial commodities, there is more information that should be introduced at the initial stage of the analysis. Moreover, if we study the behavior of many, say 15, macroeconomic time series, we need some *a priori* information on the covariance structure of the variables at the initial stage. Thus, the sequence of tests mentioned above may not be optimal for such models, since more prior information can be introduced at the initial stage.

Next, the application of statistical techniques used in the general to specific modeling approach requires some comment. The issue of model acceptance and model rejection is treated in an extensive way in Mizon's paper. If the data are in accordance with the *a priori* information specified in a model, uncertainty is reduced and in some cases the complexity of the initial general model may also be reduced. This is simplification-learning and Bayesian estimation techniques are logically very suitable in this step; see e.g. Leamer (1978). The operational possibilities of Bayesian procedures are briefly discussed in the following.

If the data are in conflict with *a priori* ideas, we have error-learning. A formal statistical technique of error-learning is not a trivial matter. One may use the Popperian view along with classical Neyman–Pearson testing techniques and only falsify models. Since models are simplifications of reality and assuming that data are reality, it can only be concluded that with an infinite amount of data (asymptotically) all models are rejected and all estimators are inconsistent. This is well known; estimates and test functions in econometrics are interpreted in a nonclassical way. That is. one compares the outcome of a test with *a priori* knowledge, e.g., on the "correct" sign of a variable or the "plausibility" of a forecast, or one is "more satisfied the larger the absolute value of the t-test is" or one adjusts the significance level of a test when the sample size grows. Thus, in practice, there is an averaging of implicit prior and model information.[1,2]

[1] At a conference on Bayesian statistics in Valencia in 1987, I remarked that Hendry's plea for test, test, test... is in practice usually changed into: average, average, average... (or integrate, integrate, integrate). Practitioners in the nonexperimental sciences in many cases

Also, for the purpose of decision analysis, a model has to be accepted, which is different from not rejecting one. The intellectual task of formulating implicit prior ideas explicitly in terms of probabilities is nontrivial. To use some Bayesian jargon, it is intellectually impossible to have all the details of a multivariate model available *a priori* in a conjugate way (that is, in a way that is easily combined with the likelihood information). Prior specification is stochastic model specification, which is not an easy topic. Some vague prior notions on some parts of the model and some very specific information on other parts of the model have to be translated into one model specification. Clearly, when using Bayesian analysis, one may also end up with a conflict between prior and data. Revision of probabilities due to errors is not trivial. More generally, sequential error-learning in terms of formal probabilities is an intrinsically different area in the sense that one cannot assign *a priori* a probability to a part of the state (or parameter) space that has initially been omitted. Indication of the uncertainty of forecasts appears to be a more operational approach.

Mizon and, earlier, Hendry (1983) avoid formal specification of prior probabilities and list a number of criteria that a model has to satisfy. In Mizon's paper, model *congruence* is stressed, which means that a model should be in accordance with *a priori* information, *data* information and properties of the *measurement* system and that the model should be able to explain the results of rival models. This last point is labeled the *encompassing* property.

One might think that the encompassing test is the "ultimate thing". Yet in the artificial example in Section II of Mizon's paper, this test is not necessary for policy analysis since the explanatory variables are generated independently and the possible effect of each explanatory variable on the dependent variable can be studied separately. The occurrence of orthogonal explanatory variables, however, is not a standard state of affairs in economics, as already emphasized at the beginning of this comment. The absence of orthogonality in the empirical section (IV) of Mizon's paper points to a problem in interpreting the estimation results of the one-but-last equation. It is suggested that the effect of a change in one explanatory variable (*dp*) on the dependent variable is measured by the regression coefficient. Here one should be aware of the condition that this holds only *ceteris paribus*. Since the explanatory variables are correlated, the total

take an implicit weighted average of their test results and their *a priori* ideas. From a scientific point of view one would like to see these different weights specified in an explicit way.

[2] Mechanical application of a classical unit root test gives implausible results, as illustrated in Mizon's paper where a bounded variable such as the unemployment percentage is "apparently" nonstationary and policy behavior with respect to changes in tax rates may also be a nonstationary process.

effect of a change in one variable on the dependent variable may be different from the value of the regression coefficient.

Even though the encompassing test is attractive in the sense that the relative probabilities of models are compared (albeit in an implicit way), it is of interest for *robustness* to know what steps have been taken in the modeling exercise using PC-GIVE; see also Pagan (1987). PC-GIVE is a menu-driven program that may be interpreted as an experts' system. At the end of the analysis, the expert using PC-GIVE achieves a ranking of the available set of models. It is of interest to have an idea of the distribution of this ranking. One approach is to apply multi-criteria analysis, where scoring rules are used to evaluate the performance of models. A prerequisite is an indication of the weights given to criteria such as congruence, data coherence, theory consistency, valid marginalization, exogeneity, etc. and an indication of the weight given to each of the models under each criterion. The simplest scoring rule is to give each model one point if it satisfies a criterion. Further, in reporting results, it is of interest to know how much of the end result of an analysis using PC-GIVE is due to "formal" procedures and how much is due to the analyst who interprets the test results.

Instrumental variable estimation plays a major role in PC-GIVE. The sensitivity of instrumental variable estimates with respect to the choice of the variables and the potential effect of such a choice on fit and inference on e.g. exogeneity are points that are relevant for practitioners. These may be illustrated in the context of the classic meat market model estimated by Tintner (1952, pp. 168–84). The demand equation for meat is given as

$$q_t = \alpha + \beta p_t + \gamma Y_t + u_t \qquad t = 1919, \ldots, 1941$$

where q_t is the consumption of meat per capita (in pounds); p_t is the price of meat; and Y_t is disposable income per capita. The supply equation relates q_t to p_t and the costs of processing meat, Z_{1t}. The model is exactly identified. The results are reported in Table 1. Column (1) contains the instrumental variable (IV) estimates using Z_{1t} as an instrument for p_t. The parameter estimates are 2SLS estimates. Note that the measure of fit may be negative in IV-estimation. IV aims for consistent estimators and not maximization of fit. Column (2) shows results for a respecified demand equation with a lagged endogenous variable with parameter δ. Parameter estimates are sensitive, the fit is worse and exogeneity is rejected.[3] In columns (1)′ and (2)′ results are reported using average hourly earnings Z_2

[3] The exogeneity test is computed as follows. Regress p_t on all instruments and add the residuals as "artificial regressors" to the demand equation. Perform a second-stage regression and use the t-value of the artificial regressor with coefficient η as a test for exogeneity; see e.g. Holly (1982). Note that the R^2 in columns (2) and (2)′ refers to the fit without the artificial regressor.

Table 1. *Different instrumental variable estimates of the demand equation parameters in Tintner's meat market model*

	Instrumental variables			
	IV1 (c, Y, Z_1)		IV2 (c, Y, Z_1, Z_2, Z_3)	
Parameter	(1)	(2)	(1)′	(2)′
α	245.5	318.5	207.5	128.6
β	−2.36	−2.88	−1.47	−0.88
γ	0.28	0.34	0.19	0.12
δ	—	−0.32	—	0.34
η	—	2.01	—	−0.06
		(2.22)*		(0.15)*
R^2	−0.04	−0.76	0.68	0.82

*Numbers in parentheses are absolute t-values. Estimation period 1920–41.

and costs of agricultural production Z_3 as instrumental variables. The fit of both the original and the respecified model is improved, while exogeneity of the price in the demand equation cannot be rejected. A general to specific modeling strategy where the information set is spelled out in the beginning of the analysis will avoid bringing in extra instruments. The example illustrates, however, that one "needs a good story" for instrumental variable choice. Bringing in good explanatory instrument variables through the back door of the consistency argument may yield desired results. However, the front door of relevant prior ideas is the proper road for achieving this. Moreover, incomplete model analysis may yield sensitive inference with respect to exogeneity, as shown in the example.

References

Fisher, F. M.: *The Identification Problem in Econometrics.* McGraw-Hill, New York, 1966.

Hendry, D. F.: Econometric modelling. The consumption function in retrospect. *Scottish Journal of Political Economy 30*, 193–220, 1983.

Hendry, D. F. & Mizon, G. E.: Evaluating dynamic econometric models by encompassing the VAR. Unpublished manuscript, Dec. 1989.

Holly, A.: A remark on Hausman's specification test. *Econometrica 50*, 749–59, 1982.

Leamer, E. E.: *Specification Searches*, Wiley, New York, 1978.

Monfort, A. & Rabemananjara, R.: From a VAR to a structural model, with an application to the wage price spiral. *Journal of Applied Econometrics 5*, 203–27, 1990.

Pagan, A.: Three econometric methodologies: A critical appraisal. *Journal of Economic Surveys 1*, 3–25, 1987.

Tintner, G.: *Econometrics.* Wiley, New York, 1952.