Journal of Econometrics 44 (1990) 67–86. North-Holland

HOW TO LIVE WITH MISSPECIFICATION IF YOU MUST

Esfandiar MAASOUMI*

Southern Methodist University, Dallas, TX 75275, USA

Statistical models are necessarily parsimonious expressions of more complicated and extensive behavioral models. Misspecification of these models is therefore endemic and inevitable. Omission of relevant variables, inclusion of 'irrelevant variables', incorrect functional forms, incompleteness of systems of relations, and incorrect distributional assumptions are both common and present simultaneously. This paper looks at the effects of several of these problems on the forecasting equations of econometric systems. It is argued that such systems can be analyzed statistically and in a way that formally admits misspecification as well as the uncertainty that surrounds its precise nature and magnitude. It turns out that the Generic Reduced Form (GRF) model and estimation techniques provide an attractive and natural way of living with this real modelling situation. Results of simulation experiments are reported which compare the proposed method with six traditional techniques.

1. Introduction

Within the context of linear simultaneous systems of econometric relations, this paper considers the effects of several categories of model misspecification and clarifies their consequences for data analysis and forecasting. It is shown that these misspecifications are the result of incorrect ('structural') reparameterizations of the basic data-generating process which is the reduced form of such systems. The traditional relations between the 'structural' coefficients and those of the reduced form are shown to be violated in several ways. Since all such misspecifications can occur simultaneously and to an extent that is not precisely known *a priori*, it is suggested that a formal and subjective model of the said violations is called for. While the increasing use of existing diagnostic and other tests is a welcome development, the proposed approach is justified since existing statistical tests are only asymptotically valid and have as yet unknown properties when several common *misspecifications are present simultaneously*. In our approach one must formally express *a priori* positions which

*I wish to thank the editor, D.J. Aigner, two anonymous referees, P. Ruud, and A. Spanos for helpful comments and discussion of a previous version. My thanks to J.-H. Jeong for computing the simulation examples and allowing me to use them. This paper was prepared for Camp Econometrics II held in March 1988 at Lake Arrowhead, CA, and is dedicated to Ted. W. Anderson.

are held over specific or *nonspecific alternative models*. Since it is under these alternatives that the estimated model is rendered misspecified, our approach amounts to specifying often vague priors on the degree and/or the form of the corresponding 'misspecification terms'. In particular, additional and often implicit prior information about the parameters of the 'estimated model' can be formally incorporated. In some cases the proposed method is an extension of the traditional Bayesian model.

A *reduced-form model* which suggests itself turns out to be formally identical to the Generic Reduced Form (GRF) model proposed earlier by the author. Direct transformations can demonstrate the formal equivalence of the nonlinear hierarchical (Bayesian) data-generating model here and in Maasoumi (1986). In this model it will generally be assumed that the unrestricted reduced form is a satisfactory *statistical* model of the data at hand, but it is subjected to *a priori* (theoretical) restrictions which cannot all be held with certainty.

The paper is organized as follows. Section 2 introduces the basic notation and assumptions of the *estimated* model and its structural reparameterization. Sections 3.1–3.3 detail three cases of common misspecifications, and section 4 contains a description of a GRF technique available for a formal treatment of these cases *when we do not know exactly which case (alternative) holds*. Section 5 provides a brief report on some simulation experiments, and section 6 concludes.

2. The estimated model

The following multivariate system of relations is under consideration between n dependent variables (Y_t), and m nonstochastic exogenous variables (Z_t), both observed at points $t \in [1, T]$:

$$Y_t = PZ_t + V_t. \tag{1}$$

V_t is the n-element vector of serially independent disturbances with zero means and a positive definite covariance matrix denoted by Ω_v. In general we assume that (1) is a satisfactory statistical model of the data. Theoretical considerations, however, may make it necessary to consider reparameterizations of (1) that represent 'structural models' of interest in an economic or other context. The latter models are given as follows:

$$BY_t + CZ_t = U_t = AX_t, \quad \text{say}, \tag{2}$$

where U_t is the 'disturbance' vector, $A = (B, C)$ is the $n \times (n+m)$ matrix of coefficients of interest about which some nonsample information is usually provided by the underlying theory, and $X_t = (Y_t', Z_t')'$ is the tth row of the observed data matrix $X = (Y\,Z)$. (2) is generally the detailed theoretical model and may be misspecified in several ways.

Assumption 1. The theoretical 'overidentifying restrictions' on A are of the simple exclusion (zero-order) type represented as follows:

$$s - S\gamma = \operatorname{vec} A. \tag{3}$$

Vec denotes stacking of matrices by rows, γ is the vector of unrestricted and unknown coefficients in A, s is a selection vector of ones and zeros which represents the normalization restrictions ($B_{ii} = 1$, all $i = 1,\dots,n$), and S is a block-diagonal selection matrix such that

$$(I \otimes X)S = X^* = \operatorname{diag}(X_1, X_2, \dots, X_n),$$

where X_i denotes those columns of X that appear as the explanatory variables in the ith equation of (2). Whenever model (2) is a 'correct' representation of the structural relations, we would refer to (1) as its reduced form and assert that

$$BP + C = 0. \tag{4}$$

These and other relations such as (3) impose nonlinear restrictions on P. But as we will see in the next section, if (2) is a misspecification of the structural relations, then (4) will not hold exactly and will be violated according to the specific type of alternative model considered.

3. Three cases of common misspecification

3.1. 'Classical' misspecification[1]

Suppose that (2) contains the correct set of variables X and their classification into endogenous and exogenous. Based on theoretical and identification considerations, however, erroneous omission (inclusion) of variables from (in) some or all of the X_i may lead to 'theoretical' misspecification of various equations. Let the correct parameterization (structural model) have coefficients B_0 and C_0. We may then write

$$B_0 = B - B_1 \quad \text{and} \quad C_0 = C - C_1. \tag{5}$$

[1] Throughout this paper we have used identical characters to denote the coefficient of the same variable in both the correct and the misspecified models. This was done only to save on notation.

Also, (1) is the reduced form of the correct model if

$$B_0 P + C_0 = 0. \tag{6}$$

It follows from these relations that

$$BP + C = B_1 P + C_1 = D_1, \quad \text{say}. \tag{7}$$

In cases where (B_0, C_0) are identified *and* (B, C) are also identified (when $B_1 = 0 = C_1$), a statistical test of $D_1 = 0$ may be useful *if no other types of misspecification are present*. Clearly, the traditional practice of imposing a degenerate prior pdf on $D_1 = 0$ is unrealistic. We do not know that $D_1 = 0$ with certainty. Some aspects of this situation have been studied by Morimune and Sawa (1980) who developed small-σ approximate properties of structural estimators with misspecifications described in (5) and $D_1 \neq 0$.

3.2. Perceived and approximate linearity of structural relations

Consider the following reduced-form relations:

$$Y_{1t} = \pi_1 + V_{1t}, \qquad Y_{2t} = \pi_2 + V_{2t}. \tag{8}$$

If the 'correct' structural parameterization is the following nonlinear model:

$$\ln Y_{1t} + \alpha = U_{1t}, \qquad Y_{2t} + \beta Y_{1t} = U_{2t}, \tag{9}$$

the implied correct restrictions on π_1 and π_2 are as follows:

$$\pi_1 = \mathrm{E}(\mathrm{e}^{-\alpha}\varepsilon_{1t}), \qquad \pi_2 = -\beta \mathrm{e}^{-\alpha}\mathrm{E}(\varepsilon_{1t}), \qquad \varepsilon_{1t} = \mathrm{e}^{U_{1t}}. \tag{10}$$

The error terms are defined conformably. For example, $V_{1t} = \mathrm{e}^{-\alpha}(\varepsilon_{1t} - \mathrm{E}(\varepsilon_{1t}))$, which has zero mean and finite variance if $\mathrm{E}(\varepsilon_{1t}^2)$ is finite.[2]

Suppose now that the following linear structural model is a misspecification of the 'true' relations in (9):

$$Y_{1t} + \alpha = \omega_{1t}, \qquad Y_{2t} + \beta Y_{1t} = \omega_{2t}. \tag{11}$$

Given that (8) is the correct data-generating model, the following incorrect restrictions on π_i are implied by the structure in (11):

$$\pi_1 + \alpha = D_{21}, \qquad \pi_2 + \pi_1 \beta = D_{22}, \tag{12}$$

$$D_2 = \begin{bmatrix} D_{21} \\ D_{22} \end{bmatrix} = 0, \qquad V_{1t} = \omega_{1t}, \qquad V_{2t} = \omega_{2t} - \beta\omega_{1t}. \tag{13}$$

[2] Other definitions of π_i are possible of course. Some of these in fact allow for both random coefficients and time dependence. The V_{it} must be defined accordingly.

By employing the usual expansion of $e^{-\alpha}$ and $e^{U_{1t}}$ in (10), however, we can easily see that $D_2 \neq 0$. The terms in D_2 are functions of α, β, and $E(\varepsilon_{1t})$. Any knowledge of these parameters may help our understanding and treatment of D_2, at least when no other types of misspecification are present.

3.3. *Perceived completeness of models*

Practical considerations as well as incomplete theoretical development usually lead to the estimation and analysis of econometric models which are partially specified. Several misspecification cases may be distinguished in such situations.

Case 1. Partial models: At least some endogenous variables are omitted from some equations, and equations which might have explained these endogenous variables are excluded from the model. Such cases of perceived completeness arise when, for instance, two parts of a model are mistakenly thought to be capable of 'decoupling', or because of *a priori* misspecification of the theoretical model. To see some of the consequences of this type of misspecification let the complete model be given as follows:[3]

$$B_0 Y_t^0 + C_0 Z_t^0 = U_t^0, \tag{14}$$

where

$$Y_t^0 = \begin{bmatrix} Y_t \\ Y_t^* \end{bmatrix}, \quad Z_t^0 = \begin{bmatrix} Z_t \\ Z_t^* \end{bmatrix}, \quad U_t^0 = \begin{bmatrix} U_t \\ U_t^* \end{bmatrix},$$
$$B_0 = \begin{bmatrix} B & B_{12} \\ B_{21} & B_2 \end{bmatrix} \quad \text{and} \quad C_0 = \begin{bmatrix} C & C_{12} \\ C_{21} & C_2 \end{bmatrix}. \tag{15}$$

The complete model in (14) implies that the model in (2) is misspecified both because it omits $B_{12}Y_t^*$ from its equations and since the equations explaining the elements of Y_t^* are missing. Solving (14) for its reduced form we find

$$Y_t^0 = P_0 Z_t^0 + V_t^0, \quad P_0 = -B_0^{-1} C_0 = \begin{bmatrix} P & P_{12} \\ P_{21} & P_2 \end{bmatrix}, \quad \text{say},$$

and

$$V_t^0 = B_0^{-1} U_t^0.$$

[3]We assume these relations are not so specified as to violate the endogeneity–exogeneity classifications derived in constructing the statistical model (1). Also, in the later sections which regard (1) as a well-specified complete statistical model, it would be convenient to consider Y_t^* as a subset of Y_t.

Once again, it follows that relations (4) are violated since now

$$BP + C = -B_{12}P_{21} = D_3 \neq 0. \tag{16}$$

Clearly model (1) will not be the correct reduced form for the partial model in (2). The consequences for the restricted estimates of P and, correspondingly, for prediction will depend on the magnitude of the errors. If these errors are the only cause of misspecification and if they are of 'small' orders of magnitude in the sample size, as Fisher (1961) considered in a different context, the effect on estimation consistency may be negligible. Of course *a priori* views about the elements of D_3 are an essential part of model building. But while it may be reasonable *a priori* to believe $D_3 = -B_{12}P_{21}$ to be negligible (or even $= 0$), this quantity is not necessarily negligible *a posteriori*, a fact that can seriously affect the purely objective inferences. This suggests that the traditional methods are unreasonable representations of the actual modelling process when they take $D_3 = 0$ with probability 1 or when they effectively allow for infinite uncertainty (variance of D_3 = infinity). A middle ground seems more generally plausible.

Inferences about the structural coefficients B, C, and Σ, the covariance matrix of the U_t, may also remain valid despite the incompleteness. Strong *a priori* decoupling conditions for this are discussed by Richard (1980, particularly theorem 3.1). These conditions include normality, $B_{12} = 0 = \Sigma_{12}$, where Σ_{12} is the covariance between Y_t and Y_t^*, and the nonexistence of any cross-restrictions between the two parts of the model. Similar results can be discerned from Barndorff-Nielsen (1978). Statistical verification of these conditions is, however, very challenging.

Case 2. 'Incomplete' models: This situation is similar to case 1 except that Y_t^* is not omitted from the model and is known to be endogenous. The perceived structural model is then given by

$$BY_t + B_{12}Y_t^* + CZ_t = U_t. \tag{17}$$

The equations for Y_t^* are not specified either because of practical difficulties or a lack of theory/interest. But it is commonly implicitly assumed that Y_t^* is also dependent on Z_t alone. If this is not so, as is probable given the lack of an articulated theory about Y_t^*, both (17) *and* the reduced-form equations for Y_t will be generally affected. Once again, working with the model in (14)–(15) we can conclude that (16) is the true parameterization relations. Here, however, we know for certain that B_{12} is not zero, and $P_{21} = 0$ will be a contradiction of the usual procedures for estimating incomplete models. Thus the correct specification would be $D_3 \neq 0$ with probability 1. Interestingly, this part of the argument does not depend on the existence of omitted predetermined variables Z_t^*, although such omission will adversely affect prediction of all the elements of Y_t^0.

Case 3. Misclassification of variables: Suppose now that model (17) is estimated based on the mistaken assumption that Y_t^* is predetermined. In this case the imposed reparameterizations are

$$B(P\colon P_{12}) + (C\colon B_{12}) = 0, \tag{18}$$

which includes $BP + C = 0$, whereas the corresponding correct relations are (16) and, if C_{12} and C_{21} are nonzero,

$$BP_{12} + B_{12}P_{21} + C_{12} = 0. \tag{19}$$

This kind of misspecification can affect our consideration of an appropriate reduced-form (statistical) model. It may lead us to consider only the joint distributions which are conditioned on Y_t^* as well as the other genuinely predetermined variables. A 'broader' statistical model such as in (14) will be conducive to tests of exogeneity and a potential avoidance of the above misclassification. But no test provides *certain* inferences, and there is room for statistical misspecification and reasonable doubt. Within the framework discussed in Spanos (1986) the above considerations argue for greater focus on the intricate interplay between theoretical and statistical model construction.

The particular example developed in section 4 will have to be extended in order to be applicable to the problems discussed in cases 2–3.

3.4. Discussion

The categories of misspecification described above are but examples of many others among which the misspecification of distributions deserves special mention. As in White (1982), one needs to study cases where the degree of misspecification is negligible, at least asymptotically and in relevant directions. But it is imperative also to develop methods of modelling and inference that are in some sense 'robust', and thus useful in the probably more common cases of nonnegligible misspecification.

It is good practice to assume that all the three misspecifications discussed in this paper are simultaneously present to some degree. One should be reluctant to assume away any of these problems and their magnitudes on *a priori* grounds, as is routinely done in the traditional approach to system estimation. Nor are the operational aspects of 'encompassing' free from such *a priori* untenable assumptions. The otherwise improved medium for the discussion of model performance afforded by the requirement of encompassing is operational in a statistical sense only when we determine with certainty the 'pseudo-true' parameters, the distributional properties of estimators, *and* the corresponding test statistics under each alternative model. As is often stated, the *raison d'etre* of encompassing is that all models are misspecified [see Hendry and Richard (1987)]. But derivation of the limiting distributions of statistics in misspecified models (even when they exist) is rather difficult [see Maasoumi and Phillips (1982)], especially when several misspecifications are allowed.

It is the case that we commonly have subjective views of how badly misspecified our models are and how well we can statistically detect and correct for these misspecifications. In the particular example developed in section 4, I assume that the statistical relations of the 'reduced-form' model in (1) have good encompassing credentials and acceptable exogeneity classifications. We can even assume that the object of theoretical interest, the 'structural' model (2), has also been obtained in the same manner. We may have successfully completed a search such as the one proposed by H. White in this volume. Given these assumptions, it may be reasonable to believe *a priori* that $D_i = 0$, $i = 1, 2, 3$, but highly unrealistic to believe so with certainty. It would be equally unreasonable in this situation to have no confidence in such models and proceed with highly imprecise inferences that often characterize the *apparently* atheoretical analyses of 'unrestricted' versions of (1), such as in VARs. The middle ground that is supported by these considerations is a formal representation of uncertain knowledge about the likely magnitudes of D_i. This knowledge is fundamentally partly subjective and partly the effect of *informal awareness of objective statistical evidence* and the inadequacies of the data (e.g., sample size) and our techniques. It is largely an expression of odds on alternatives.

4. A more complete data-generating process

A formal representation of these 'irreducible' misspecifications can be achieved by specifying probability distributions on D_i, possibly with other probability distributions generating the parameters of this distribution and incorporating any information about these deeper parameters. Alternatively, we may study the sensitivity (robustness) of our inferences with respect to a wide range of values for these hyper-parameters. The joint likelihood of the data *and* the parameters is then the medium for inferences. This is similar to a nonlinear hierarchical model with well-known Bayesian interpretations. Specifically, for the models given in (1)–(2) and some $D_i = D$, let

$$\operatorname{vec} D \sim f(0, \Omega_d) \tag{20}$$

denote the density function generating D. Also, let $\bar{P}$ denote the reduced-form coefficients of the 'perceived' structural model (2). From the factual relations

$BP + C = D$ we see that

$$P = -B^{-1}C + B^{-1}D = \bar{P} + e, \quad \text{say}. \tag{21}$$

The implied probability distribution for P is derived from (20–(21). Note that $\bar{P}$ will contain all the theoretical restrictions of interest on $A = (B\ C)$, for example those in (3). We may also specify a nonzero mean in (20) and use (e.g.) noninformative or conjugate 'prior' distributions on the mean, the variance (Ω_d), and other parameters of $f(\cdot)$.

To construct the *joint likelihood* for the data *and* P, we use the following relations for the model in this paper:

$$\Omega_p = (B^{-1} \otimes I)\Omega_d(B'^{-1} \otimes I) \quad \text{and} \quad (Q' \otimes I)(s - S\gamma) = \text{vec}\, D, \tag{22}$$

where

$$Q = \begin{bmatrix} P \\ I \end{bmatrix},$$

and the second equation in (22) is a row vectorization of $BP + C = D$. It is clear that this model is formally identical with the Generic Reduced Form (GRF) model proposed in Maasoumi (1986). As is readily discernible from the likelihood expressions (9)–(10) of that paper, the *joint* pseudo-likelihood estimators of P and γ are identical with the *joint likelihood* estimators under normality assumptions for both the data and the parameter distributions. They can also be interpreted as the joint posterior *mode* under the same distributional assumptions and the same subjective information. The optimization problem is one of maximizing the following function:

$$\begin{aligned} &\frac{T}{2}\log|\Omega_v|^{-1} - \tfrac{1}{2}\operatorname{tr}\Omega_v^{-1}Y'Q_ZY \\ &- \tfrac{1}{2}\operatorname{tr}\Omega_v^{-1}(P_{\text{uls}} - P)(Z'Z)(P_{\text{uls}} - P)' + \tfrac{1}{2}\log|\Omega_d|^{-1} \\ &- \tfrac{1}{2}(s - S\gamma)'(I \otimes Q)\Omega_d^{-1}(I \otimes Q')(s - S\gamma), \end{aligned} \tag{23}$$

where Ω_v is the covariance matrix of the serially independent errors $V_t = (V_{1t}, V_{2t})$, $Q_Z = I - Z(Z'Z)^{-1}Z'$, and P_{uls} is the unrestricted LS estimate of P in (1). Given our discussion in section 3.4, it is generally reasonable to expect the latter to be consistent but not efficient. This will be so even if the structural relations in (2) are misspecified in the manner of section 3.1.

The solution of the above optimization problem has the familiar form and properties of a Bayes estimate. A linear approximation of the optimand (or the first-order conditions) was worked out in Maasoumi (1986) which, for suitable weight matrices W, leads to the following weighted average GRF estimator P^*:

$$\operatorname{vec} P^* = W \operatorname{vec} P_{\text{uls}} + (I - W)\operatorname{vec} P^+, \tag{24}$$

where P^+ is a restricted (derived) estimator of P such as FIML or 3SLS. The properties of P^* in both small- and large-sample cases have been studied in Maasoumi (1986) and Maasoumi and Jeong (1988). In particular, under the 'classical normal' assumptions, P^* has some finite moments where other traditional reduced-form estimators (except FIML) do not [see Sargan (1988)]. The consequences of this type of generic data-generation model for structural estimation (of γ) were also studied in Maasoumi (1986). Some simulation evidence is presented in the next section.

5. Monte Carlo simulations

A brief simulation example based on 200 replications is described here to demonstrate some of the desirable properties of the GRF estimator of P in two small models and four experimental settings in each case.[4] Table 1 describes the salient features of the experiments. The last experiment uses the data of model B to estimate model A. This means that there are problems of included and omitted variables in, respectively, the first and second equations of model A. The situation for model B is exactly reversed. The implied values for the discrepancy matrix D were computed according to expression (7) of section 3.1. For models A and B these are, respectively,

$$D = \begin{bmatrix} 0 & 0 & -4 & 0 \\ 0 & 0 & 4 & 0 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 0 & 0 & 4 & 0 \\ 0 & 0 & -4 & 0 \end{bmatrix}.$$

In order to mimic reality, our prior distributions on these matrices are centered at zero means with empirical variances computed as in Maasoumi (1986).

The change in the concentration matrices between experiments I and III, and the way we have generated the strongly exogenous variables Z, ensures a

[4] A detailed description of these experiments and some other empirical applications appear in Maasoumi and Jeong (1988).

Table 1

Two experimental models.

	Model A	Model B
(a) $A = [B : \Gamma] =$	$\begin{bmatrix} 1 & -0.5 & : & 0.5 & -0.75 & 0 & 0 \\ -4 & 1 & : & 0 & 0 & 4.0 & -1.6 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.5 & : & 0.5 & -0.75 & 4.0 & 0 \\ -4 & 1 & : & 0 & 0 & 0 & -1.6 \end{bmatrix}$
Σ and Ω	$\begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}, \begin{bmatrix} 1.75 & 6.0 \\ 6.0 & 21.0 \end{bmatrix}$	Same as model A
(b) $P =$	$\begin{bmatrix} 0.5 & -0.75 & 2.0 & -0.8 \\ 2.0 & -3.0 & 4.0 & -1.6 \end{bmatrix}$	$\begin{bmatrix} 0.5 & -0.75 & 4.0 & -0.8 \\ 2.0 & -3.0 & 16.0 & -1.6 \end{bmatrix}$
	Experiment I ($T = 24$, higher centrality, correct specification)	
(a) $\lvert Z'Z \rvert$	$\lvert Z'Z \rvert = \lvert 24I_4 \rvert = 24^4$	Same as model A
(b) Degree of overidentification: $v = v_1 + v_2$	$2 = 1 + 1$	$2 = 0 + 2$
(c) Noncentrality parameter: $P(Z'Z/T)P'$	$\begin{bmatrix} 5.45 & 12.53 \\ 12.53 & 31.56 \end{bmatrix}$	$\begin{bmatrix} 17.45 & 68.54 \\ 68.54 & 271.59 \end{bmatrix}$
(d) $\mathrm{tr}(\Omega^{-1}(PZ'ZP'))/T$	25.83	25.83
	Experiment II ($T = 48$, higher centrality, correct specification)	
(a) $\lvert Z'Z \rvert$,	$\lvert Z'Z \rvert = \lvert 48I_4 \rvert = 48^4$	Same as model A
	Experiment III ($T = 24$, lower centrality, correct specification)	
(a) $\lvert Z'Z \rvert$	$\lvert Z'Z \rvert = \lvert I_4 \rvert = 1.0$	Same as model A
(c) $PZ'ZP'/T$	$\begin{bmatrix} 0.22 & 0.52 \\ 0.52 & 1.32 \end{bmatrix}$	$\begin{bmatrix} 0.73 & 2.86 \\ 2.86 & 11.30 \end{bmatrix}$
(d) $\mathrm{tr}(\Omega^{-1}(P'Z'ZP))/T$	1.076	1.076
	Experiment IV ($T = 24$, higher centrality, misspecification)	
Perceived structural model	Model B	Model A
Exogenous variables are based on		

$$Z' = l_t' \otimes \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ \sqrt{2} & 0 & -\sqrt{2} & 0 & \sqrt{2} & 0 & -\sqrt{2} & 0 \\ 0 & \sqrt{2} & 0 & -\sqrt{2} & 0 & \sqrt{2} & 0 & -\sqrt{2} \end{bmatrix} = l_t' \otimes Z^{*\prime}$$

l_i = i-element vector of ones, $\bar{Z}_j^* = 0$, $j = 1, \ldots, 4$, and $Z^{*\prime}Z^* = TI_4$

Table 2

Distribution of the reduced-form coefficient estimates of Z (true value = 0.500).

	Mean	Std. dev.	Skewness	Kurtosis	Median	Length of interval 50%	80%	100%
Experiment I. Model A, correct specification, sample size 24								
ULS	0.483	0.270	0.36	−0.04	0.47	0.39	0.69	1.58
DRF-2SLS	0.504	0.230	0.63	0.64	0.48	0.32	0.59	1.35
DRF-3SLS	0.520	0.216	0.59	0.92	0.52	0.30	0.51	1.43
DRF-FIML	0.501	0.216	0.67	1.14	0.48	0.30	0.50	1.50
PRRF	0.483	0.270	0.36	−0.04	0.47	0.39	0.69	1.58
MSRF	0.515	0.221	0.54	0.80	0.51	0.30	0.52	1.43
GRF ($c^* = 1.0$)	0.499	0.221	0.66	0.84	0.47	0.31	0.53	1.39
GRF ($c^* = 0.05$)	0.500	0.218	0.68	0.83	0.48	0.31	0.52	1.30
Experiment I. Model B, correct specification, sample size 24								
ULS	0.483	0.270	0.36	−0.04	0.47	0.39	0.69	1.58
DRF-2SLS	0.487	0.201	0.30	−0.07	0.48	0.26	0.53	1.07
DRF-3SLS	0.491	0.173	0.14	0.38	0.50	0.25	0.43	1.16
DRF-FIML	0.488	0.171	0.01	0.54	0.49	0.25	0.42	1.20
PRRF	0.483	0.270	0.36	−0.04	0.47	0.39	0.69	1.58
MSRF	0.491	0.173	0.14	0.38	0.50	0.25	0.43	1.16
GRF ($c^* = 1.0$)	0.487	0.177	0.18	0.06	0.49	0.26	0.45	1.08
GRF ($c^* = 0.05$)	0.488	0.172	0.03	0.51	0.49	0.25	0.42	1.19
Experiment II. Model A, correct specification, sample size 48								
ULS	0.477	0.185	0.14	−0.16	0.46	0.22	0.51	0.94
DRF-2SLS	0.492	0.154	0.38	−0.13	0.47	0.20	0.41	0.79
DRF-3SLS	0.506	0.145	0.45	0.10	0.49	0.19	0.37	0.77
DRF-FIML	0.494	0.147	0.34	0.29	0.48	0.18	0.37	0.82
PRRF	0.477	0.185	0.14	−0.16	0.46	0.22	0.51	0.94
MSRF	0.508	0.146	0.45	0.17	0.49	0.18	0.36	0.78
GRF ($c^* = 1.0$)	0.493	0.147	0.42	0.06	0.48	0.18	0.36	0.76
GRF ($c^* = 0.05$)	0.494	0.146	0.45	0.07	0.48	0.18	0.37	0.76
Experiment II. Model B, correct specification, sample size 48								
ULS	0.491	0.184	0.37	−0.05	0.47	0.26	0.50	1.10
DRF-2SLS	0.494	0.138	0.40	−0.11	0.48	0.17	0.37	0.75
DRF-3SLS	0.498	0.123	0.25	0.52	0.49	0.16	0.30	0.84
DRF-FIML	0.496	0.124	0.18	0.80	0.49	0.16	0.30	0.89
PRRF	0.491	0.184	0.37	−0.05	0.47	0.26	0.50	1.10
MSRF	0.498	0.123	0.25	0.52	0.49	0.16	0.30	0.84
GRF ($c^* = 1.0$)	0.496	0.124	0.30	0.36	0.49	0.16	0.31	0.79
GRF ($c^* = 0.05$)	0.496	0.124	0.19	0.75	0.49	0.16	0.30	0.89
Experiment III. Model A, correct specification, lower centrality, sample size 24								
ULS	0.415	1.325	0.36	−0.04	0.36	1.93	3.36	7.76
DRF-2SLS	0.672	7.528	11.68	176.25	0.44	2.19	3.77	139.34
DRF-3SLS	0.421	1.344	0.22	0.03	0.38	2.05	3.35	8.46
DRF-FIML	0.378	1.265	0.10	1.66	0.28	1.61	3.04	10.57
PRRF	0.415	1.325	0.36	−0.04	0.36	1.93	3.36	7.76
MSRF	0.411	1.348	0.25	0.04	0.35	1.96	3.32	8.46
GRF ($c^* = 1.0$)	0.413	1.257	0.41	0.31	0.35	1.73	3.21	7.85
GRF ($c^* = 0.05$)	0.416	1.292	0.36	0.12	0.35	1.85	3.25	7.87

Table 2 (continued)

	Mean	Std. dev.	Skewness	Kurtosis	Median	Length of interval 50%	80%	100%
Experiment III. Model B, correct specification, lower centrality, sample size 24								
ULS	0.415	1.325	0.36	−0.04	0.36	1.93	3.36	7.76
DRF-2SLS	−0.380	11.307	−15.11	243.82	0.45	1.61	3.22	182.00
DRF-3SLS	0.304	2.802	−12.61	184.72	0.51	1.41	2.54	44.62
DRF-FIML	0.436	0.866	0.25	0.75	0.43	1.08	2.13	5.64
PRRF	0.415	1.325	0.36	−0.04	0.36	1.93	3.36	7.76
MSRF	0.474	1.018	0.21	0.73	0.51	1.36	2.47	6.78
GRF ($c^* = 1.0$)	0.437	0.950	0.35	0.57	0.43	1.27	2.32	6.09
GRF ($c^* = 0.05$)	0.438	0.959	0.39	0.58	0.43	1.28	2.32	6.09
Experiment IV. Model A, misspecified as model B, sample size 24								
ULS	0.483	0.270	0.36	−0.04	0.47	0.39	0.69	1.58
DRF-2SLS	2.845	37.421	3.07	77.49	1.59	2.52	14.28	715.49
DRF-3SLS	1.314	18.733	−0.77	71.63	0.94	1.69	7.48	344.12
PRF-FIML	−0.102	0.484	−10.43	135.47	−0.09	0.20	0.41	7.34
PRRF	0.483	0.270	0.36	−0.04	0.47	0.39	0.69	1.58
MSRF	1.299	18.733	−0.77	71.62	0.82	1.70	7.48	344.12
GRF($c^* = 1.0$)	0.324	0.298	0.66	0.27	0.28	0.37	0.79	1.61
GRF ($c^* = 0.05$)	0.304	0.327	0.38	−0.21	0.27	0.45	0.89	1.80
Experiment IV. Model B, misspecified as model A, sample size 24								
ULS	0.483	0.270	0.38	−0.04	0.47	0.39	0.69	1.58
DRF-2SLS	1.053	12.655	2.75	100.39	0.52	1.66	5.07	264.10
DRF-3SLS	0.444	1.643	−14.25	217.81	0.51	0.39	0.71	27.64
PRF-FIML	0.460	0.360	0.36	1.98	0.43	0.39	0.88	2.84
PRRF	0.483	0.270	0.36	−0.04	0.47	0.39	0.69	1.58
MSRF	0.483	0.369	0.37	−0.04	0.47	0.39	0.68	1.58
GRF ($c^* = 1.0$)	0.499	0.230	0.59	0.52	0.47	0.33	0.55	1.33
GRF ($c^* = 0.05$)	0.501	0.230	0.53	0.50	0.48	0.33	0.53	1.32

wide and yet tidy coverage of *both* the sample and parameter spaces. Four well-known traditional reduced-form estimators and three 'improved' methods are compared. We report such traditional characteristics as mean, standard error, skewness, and kurtosis, as well as the median and several 'probability ranges' which are more meaningful when studying estimators without finite-sample moments (such as 2SLS and 3SLS reduced-form estimators). Also, two typical values of the GRF estimates are reported among a larger set obtained by searching over a grid of values for the prior variance matrix Ω_d (c^* denotes the common factor which was varied). The distributions of the coefficients of Z_1 are reported in table 2. These results complement the findings of Maasoumi and Jeong (1988). GRF estimates do very well even in situations that are commonly thought to be favorable to the traditional and asymptotically efficient estimators. The Modified Stein-like Reduced Form (MSRF)

Table 3

Case I. Small sample ($T = 24$), higher centrality, correct specification.

	Mean	Std. dev.	Skewness	Kurtosis	Median	Length of interval 50%	80%	100%
Model A. One-period-lead forecasts of $Y_1 = 2.578$, one degree overidentified								
ULS	2.479	0.519	−0.18	0.28	2.50	0.63	1.33	2.93
DRF-2SLS	2.562	0.457	−0.27	0.81	2.58	0.54	1.12	2.81
DRF-3SLS	2.586	0.443	−0.15	0.84	2.58	0.53	1.11	2.83
DRF-FIML	2.529	0.436	−0.05	1.04	2.53	0.52	1.08	2.95
PRRF	2.412	0.573	−0.26	0.35	2.46	0.67	1.47	3.38
MSRF	2.553	0.446	−0.11	0.70	2.56	0.55	1.16	2.84
GRF ($c^* = 1.0$)	2.517	0.440	−0.14	0.91	2.53	0.50	1.07	2.80
GRF ($c^* = 0.05$)	2.522	0.432	−0.16	0.89	2.53	0.52	1.08	2.75
Model A. One-period-lead forecasts of $Y_2 = 4.657$, one degree overidentified								
ULS	4.332	1.834	−0.22	0.45	4.43	2.19	4.76	11.14
DRF-2SLS	4.636	1.687	−0.34	0.74	4.65	1.99	4.05	10.13
DRF-3SLS	4.704	1.631	−0.21	0.77	4.69	1.93	4.06	10.21
DRF-FIML	4.491	1.612	−0.15	0.84	4.47	1.99	3.98	9.89
PRRF	4.305	1.936	−0.22	0.22	4.39	2.32	5.07	11.16
MSRF	4.584	1.640	−0.19	0.62	4.64	1.94	4.29	10.27
GRF ($c^* = 1.0$)	4.451	1.619	−0.23	0.81	4.46	1.89	4.03	9.97
GRF ($c^* = 0.05$)	4.463	1.593	−0.25	0.78	4.43	1.90	4.00	9.86
Model B. One-period-lead forecasts of $Y_1 = 5.407$, just identified								
ULS	5.308	0.519	−0.18	0.28	5.33	0.63	2.50	2.93
DRF-2SLS	5.339	0.461	−0.09	0.71	5.35	0.55	1.33	2.86
DRF-3SLS	5.372	0.447	0.04	0.73	5.37	0.49	1.14	2.76
DRF-FIML	5.352	0.446	0.04	0.76	5.36	0.50	1.14	2.77
PRRF	5.308	0.519	−0.18	0.28	5.33	0.63	1.11	2.93
MSRF	5.372	0.447	0.04	0.73	5.37	0.49	1.33	2.76
GRF ($c^* = 1.0$)	5.345	0.447	0.01	0.70	5.36	0.51	1.14	2.80
GRF ($c^* = 0.05$)	5.353	0.447	0.05	0.76	5.36	0.51	1.14	2.77
Model B. One-period-lead forecast of $Y_2 = 21.629$, two degrees overidentified								
ULS	21.304	1.834	−0.22	0.45	21.41	2.19	4.76	11.14
DRF-2SLS	21.358	1.733	−0.20	0.71	21.45	1.99	4.36	10.99
DRF-3SLS	21.490	1.674	−0.08	0.74	21.54	1.72	4.42	10.59
DRF-FIML	21.434	1.674	−0.07	0.77	21.44	1.73	4.31	10.60
PRRF	21.231	1.978	−0.26	0.26	21.34	2.54	5.13	11.25
MSRF	21.490	1.674	−0.08	0.74	21.54	1.72	4.42	10.59
GRF ($c^* = 1.0$)	21.413	1.671	−0.10	0.74	21.48	1.75	4.31	10.69
GRF ($c^* = 0.05$)	21.439	1.675	−0.07	0.78	21.46	1.71	4.28	10.60

Table 4

Case II. Medium sample ($T = 48$), higher centrality, correct specification.

	Mean	Std. dev.	Skewness	Kurtosis	Median	Length of interval 50%	80%	100%
Model A. One-period-lead forecasts of Y_1 = 2.578, one degree overidentified								
ULS	2.501	0.346	−0.01	−0.13	2.51	0.46	0.93	1.92
DRF-2SLS	2.548	0.304	0.03	0.03	2.55	0.40	0.79	1.62
DRF-3SLS	2.568	0.296	0.03	0.18	2.57	0.40	0.75	1.73
DRF-FIML	2.537	0.295	−0.01	0.09	2.55	0.41	0.75	1.69
PRRF	2.467	0.375	−0.18	0.21	2.48	0.47	0.97	2.20
MSRF	2.559	0.295	0.03	0.11	2.57	0.40	0.76	1.70
GRF ($c^* = 1.0$)	2.531	0.292	0.01	0.09	2.55	0.41	0.73	1.68
GRF ($c^* = 0.05$)	2.534	0.291	0.02	0.09	2.54	0.41	0.74	1.68
Model A. One-period-lead forecasts of Y_2 = 4.657, one degree overidentified								
ULS	4.408	1.214	0.07	0.07	4.46	1.53	3.15	6.76
DRF-2SLS	4.571	1.119	0.12	0.05	4.62	1.54	2.77	5.98
DRF-3SLS	4.635	1.077	0.13	0.20	4.59	1.50	2.70	6.37
DRF-FIML	4.520	1.077	0.06	0.09	4.58	1.55	2.67	6.26
PRRF	4.384	1.296	0.06	−0.11	4.44	1.72	3.31	7.27
MSRF	4.602	1.078	0.11	0.15	4.58	1.49	2.69	6.32
GRF ($c^* = 1.0$)	4.501	1.067	0.09	0.13	4.54	1.39	2.67	6.22
GRF ($c^* = 0.05$)	4.509	1.064	0.10	0.11	4.54	1.47	2.66	6.21
Model B. One-period-lead forecasts of Y_1 = 5.407, just identified								
ULS	5.369	0.344	0.17	0.20	5.36	0.43	0.89	2.17
DRF-2SLS	5.387	0.310	0.18	0.05	5.38	0.40	0.80	1.76
DRF-3SLS	5.405	0.301	0.10	0.03	5.39	0.43	0.75	1.69
DRF-FIML	5.393	0.302	0.09	0.02	5.39	0.42	0.76	1.67
PRRF	5.369	0.344	0.17	0.20	5.36	0.43	0.89	2.17
MSRF	5.405	0.301	0.10	0.03	5.39	0.43	0.75	1.69
GRF ($c^* = 1.0$)	5.390	0.300	0.14	0.06	5.38	0.41	0.76	1.71
GRF ($c^* = 0.05$)	5.394	0.302	0.09	0.02	5.59	0.42	0.76	1.67
Model B. One-period-lead forecasts of Y_2 = 21.629, two degrees overidentified								
ULS	21.490	1.210	0.22	0.36	21.40	1.56	3.07	7.61
DRF-2SLS	18.383	1.155	0.19	0.22	21.49	1.46	2.97	6.78
DRF-3SLS	21.597	1.120	0.10	0.21	21.58	1.46	2.80	6.50
DRF-FIML	21.563	1.124	0.08	0.19	21.58	1.48	2.74	6.43
PRRF	21.452	1.299	0.20	0.36	21.38	1.68	3.31	8.52
MSRF	21.597	1.120	0.10	0.21	21.58	1.46	2.80	6.50
GRF ($c^* = 1.0$)	21.554	1.115	0.13	0.24	21.51	1.38	2.83	6.54
GRF ($c^* = 0.05$)	21.564	1.123	0.09	0.19	21.58	1.45	2.76	6.44

Table 5

Case III. Small sample ($T = 24$), lower centrality, correct specification.

	Mean	Std. dev.	Skewness	Kurtosis	Median	Length of interval 50%	80%	100%
Model A. One-period-lead forecasts of $Y_1 = 0.526$, one degree overidentified								
ULS	0.427	0.519	−0.18	0.28	0.45	0.63	1.33	2.93
DRF-2SLS	4.082	44.523	14.67	222.83	0.64	0.92	2.52	706.02
DRF-3SLS	2.223	24.159	15.62	245.84	0.65	0.84	1.91	387.52
DRF-FIML	0.648	3.099	15.26	238.33	0.44	0.55	1.28	49.77
PRRF	0.321	0.557	0.017	0.003	0.36	0.70	1.46	3.04
MSRF	2.037	24.125	15.707	247.728	0.54	0.75	1.62	387.52
GRF ($c^* = 1.0$)	0.439	0.469	−0.319	0.371	0.47	0.60	1.30	2.86
GRF ($c^* = 0.05$)	0.438	0.505	−0.250	0.272	0.45	0.62	1.35	2.87
Model A. One-period-lead forecasts of $Y_2 = 0.951$, one degree overidentified								
ULS	0.626	1.834	−0.22	0.45	0.73	2.19	4.76	11.14
DRF-2SLS	12.818	147.477	14.83	227.03	1.42	3.09	8.49	2345.45
DRF-3SLS	6.748	80.421	15.62	245.90	1.46	2.83	6.39	1289.22
DRF-FIML	1.372	10.250	15.67	236.54	0.75	2.03	4.55	164.89
PRRF	0.634	1.853	−0.23	0.39	0.73	2.28	4.87	11.08
MSRF	6.024	80.294	15.72	248.06	1.00	2.55	5.69	1288.34
GRF ($c^* = 1.0$)	0.684	1.767	−0.40	0.51	0.76	2.09	4.73	10.73
GRF ($c^* = 0.05$)	0.678	1.796	−0.33	0.37	0.78	2.24	4.82	10.72
Model B. One-period-lead forecasts of $Y_1 = 1.104$, just identified								
ULS	1.004	0.519	−0.18	0.28	1.03	0.63	1.33	2.93
DRF-2SLS	0.898	2.534	−14.05	212.53	1.09	0.60	1.35	40.95
DRF-3SLS	1.094	0.744	−7.09	82.73	1.13	0.47	1.17	10.44
DRF-FIML	1.036	0.440	0.05	0.73	1.05	0.50	1.11	2.79
PRRF	1.004	0.519	−0.18	0.28	1.03	0.63	1.33	2.93
MSRF	1.132	0.461	−0.11	0.97	1.13	0.47	1.17	3.03
GRF ($c^* = 1.0$)	1.044	0.443	0.00	0.77	1.04	0.47	1.11	2.88
GRF ($c^* = 0.05$)	1.045	0.446	−0.06	0.80	1.06	0.48	1.13	2.89
Model B. One-period-lead forecasts of $Y_2 = 4.415$, two degrees overidentified								
ULS	4.090	1.834	−0.22	0.45	4.19	2.19	4.76	11.14
DRF-2SLS	3.513	10.116	−14.20	215.83	4.28	2.16	5.29	164.07
DRF-3SLS	4.289	2.908	−7.60	90.74	4.46	1.81	4.65	41.31
DRF-FIML	4.175	1.654	−0.12	0.77	4.22	1.83	4.18	10.68
PRRF	3.933	1.985	−0.29	0.25	4.03	2.50	5.11	11.20
MSRF	4.444	1.733	−0.24	0.93	4.45	1.78	4.63	11.00
GRF ($c^* = 1.0$)	4.198	1.658	−0.15	0.81	4.23	1.79	4.24	10.77
GRF ($c^* = 0.05$)	4.191	1.674	−0.20	0.81	4.22	1.77	4.26	10.70

Table 6

Case IV. Small sample ($T = 24$), higher centrality, misspecification.[a]

	Mean	Std. dev.	Skewness	Kurtosis	Median	Length of interval 50%	80%	100%
			Model A. Forecasts of $Y_1 = 2.578$, $\upsilon_1 = 1$ misspecified as $\upsilon_1 = 0$					
ULS	2.479	0.519	−0.18	0.28	2.50	0.63	1.33	2.93
DRF-2SLS	−1.027	44.819	0.14	56.15	0.44	5.09	25.67	779.69
DRF-3SLS	1.010	25.321	2.31	59.44	1.67	3.19	13.99	461.85
PRF-FIML	2.428	1.390	−0.55	0.67	2.73	1.50	3.95	8.45
PRRF	2.479	0.519	−0.18	0.28	2.50	0.63	1.33	2.93
MSRF	1.040	25.322	1.31	59.43	1.93	3.16	13.99	461.85
GRF ($c^* = 1.0$)	2.983	0.678	0.03	−0.47	3.01	1.00	1.76	3.43
GRF ($c^* = 0.05$)	2.981	0.743	0.12	4.90	2.93	1.11	1.95	3.81
			Model A. Forecasts of $Y_2 = 4.657$, $\upsilon_2 = 1$ misspecified as $\upsilon_2 = 2$					
ULS	4.332	1.834	−0.22	0.45	4.43	2.19	4.76	11.14
DRF-2SLS	−2.257	106.369	0.36	57.58	1.08	12.40	57.33	1882.57
DRF-3SLS	2.406	60.853	1.35	60.08	4.17	7.65	33.02	1115.13
DRF-FIML	4.750	4.452	−0.60	0.29	5.41	5.25	12.56	24.30
PRRF	5.855	1.390	0.17	0.31	5.80	1.62	3.57	8.00
MSRF	2.212	60.845	1.36	60.14	3.81	7.25	33.02	1115.13
GRF ($c^* = 1.0$)	6.186	2.455	−0.06	−0.48	6.37	3.52	6.31	12.39
GRF ($c^* = 0.05$)	6.153	2.686	0.08	−0.52	5.72	4.10	6.78	14.43
			Model B. Forecasts of $Y_1 = 5.407$, $\upsilon_2 = 0$ misspecified as $\upsilon_1 = 1$					
ULS	5.308	0.519	−0.18	0.28	5.33	0.63	1.33	2.93
DRF-2SLS	15.109	117.654	15.18	236.15	5.34	4.07	10.80	1906.14
DRF-3SLS	12.484	88.047	15.26	238.17	5.74	2.47	7.12	1450.06
DRF-FIML	3.501	6.103	0.40	58.28	3.88	3.35	5.68	109.49
PRRF	5.370	0.519	−0.20	0.30	5.39	0.66	1.34	2.87
MSRF	5.314	0.517	−0.17	0.26	5.34	0.62	1.37	2.94
GRF ($c^* = 1.0$)	5.328	0.707	0.16	−0.21	5.28	1.01	1.76	3.76
GRF ($c^* = 0.05$)	5.384	0.637	0.31	0.85	5.34	0.79	1.60	4.34
			Model B. Forecasts of $Y_2 = 21.629$, $\upsilon_2 = 2$ misspecified as $\upsilon_2 = 1$					
ULS	21.304	1.834	−0.22	0.45	21.41	2.19	4.76	11.14
DRF-2SLS	59.248	456.966	15.16	235.74	22.44	15.05	43.41	7404.70
DRF-3SLS	49.033	341.950	15.25	237.80	23.00	9.44	28.73	5634.29
DRF-FIML	14.662	23.727	0.40	58.59	16.23	12.92	21.39	424.91
PRRF	21.277	1.936	−0.22	0.22	21.36	2.32	5.09	11.16
MSRF	21.325	1.830	−0.22	0.43	21.45	2.16	4.79	11.19
GRF ($c^* = 1.0$)	21.363	2.332	0.11	0.01	21.27	3.09	6.09	12.90
GRF ($c^* = 0.05$)	21.558	2.154	0.25	1.00	21.46	2.73	5.62	14.79

[a] $\upsilon_1(\upsilon_2)$ = degree of overidentification of the first (second) equation.

Table 7

Ranking of forecasts.[a]

	Case I					Case II					Case III					Case IV					
	Model A		Model B			Model A		Model B			Model A		Model B			Model A		Model B			Grand
	Y_1	Y_2	Y_1	Y_2	Totals	Y_1	Y_2	Y_1	Y_2	Totals	Y_1	Y_2	Y_1	Y_2	Totals	Y_1	Y_2	Y_1	Y_2	Totals	totals
R.M.S.E. (standard deviation of samples) rankings																					
ULS	7	7	7	7	28	7	7	7	7	28	3	3	5	5	16	1	2	2	2	7	79
DRF-2SLS	6	6	6	6	24	6	6	6	6	24	8	8	8	8	32	8	8	8	8	32	112
DRF-3SLS	5	3	2	2	12	4	4	2	2	12	7	7	7	7	28	6	7	7	7	27	79
DRF-FIML	3	3	4	5	15	2	2	1	2	7	5	5	1	1	12	5	5	6	6	22	56
PRRF	8	8	7	8	31	8	8	7	8	31	4	4	5	6	19	1	1	2	3	7	72
MSRF	3	5	2	2	12	5	5	2	2	14	6	6	4	4	20	6	6	1	1	14	60
GRF ($c = 1.0$)	2	2	1	1	6	3	3	2	1	9	1	1	2	2	6	3	3	5	5	16	60
GRF ($c = 1.0$)	2	2	1	1	6	3	3	2	1	9	1	1	2	2	6	3	3	5	5	16	37
GRF ($c = 0.05$)	1	1	4	4	10	1	1	2	5	9	2	2	3	3	10	4	4	4	4	16	45
Quintile (80% range) length rankings																					
ULS	7	7	7	7	28	7	7	8	7	29	3	3	6	6	18	1	2	1	1	5	82
DRF-2SLS	6	6	6	6	24	5	4	6	4	19	8	8	8	8	32	8	8	8	8	32	107
DRF-3SLS	3	5	1	3	12	4	5	2	5	16	7	7	4	5	23	6	6	7	7	26	78
DRF-FIML	3	2	5	1	11	2	1	2	2	7	1	1	1	1	4	5	5	6	6	22	46
PRRF	8	8	7	8	31	8	8	1	8	25	5	5	6	7	23	1	1	2	3	7	86
MSRF	5	4	1	3	13	6	6	6	5	23	6	6	4	4	20	6	6	3	2	17	70
GRF ($c = 1.0$)	1	2	3	5	11	2	3	2	2	9	2	2	1	2	7	3	3	5	5	16	43
GRF ($c = 0.05$)	2	1	4	2	9	1	2	2	1	6	4	4	3	3	14	4	4	4	4	16	45

[a] The forecast-period values of exogenous variables were $(1\ 1\ \sqrt{2}\ 0)$.

estimator performs very well except in a few instances, as in the misspecified and low centrality model A, where it behaves almost as badly as 3SLS. The χ^2 test defining MSRF is either undefined due to a lack of overidentification, or occasionally fails to reject the model, thus making the estimator identical to its 3SLS component [see Maasoumi (1978)]. In the other situations MSRF does a good job of removing the outliers in the distribution of the 3SLS.

The Partially Restricted Reduced Form (PRRF) estimator, while better than the traditional methods in most cases, does not fair as well except when P_{uls} is expected to do very well, as in the exactly misspecified cases of experiment IV. P_{uls} is impressive except for a lack of relative precision with large samples and correct specification.

As for 'the acid test' of model performance, tables 3–6 report on the 'one-period-ahead' forecasting performance of these methods. Once again the two GRF examples come first overall, and do very well indeed in every experiment. There appears to be no costs in using GRF even when the traditional (derived) reduced-form (DRF) estimators may be expected to do well. FIML and 3SLS are better than 2SLS, the most widely used of the traditional methods! 2SLS is consistently the worst performer in almost every respect. Our results also show that median is a very good (robust) estimator in these models. Table 7 summarizes simple cumulative ranking of these forecasts on the basis of Root Mean Square Error and the quintile range, respectively. These summaries are clearly consistent with our observations based on tables 3–6.

6. Conclusions

It was argued that after all the statistical checks on a model have made a case for its use, reasonable doubts concerning its adequacy *and* the power of the model selection methodology persist. A proposed method was demonstrated in a common-systems context to be able to formally represent this state of affairs. This method formally incorporates otherwise implicit and subjective positions with respect to the *alternative models*. In this way our statistical analysis of the preferred model is affected by how strongly we hold it *and* its alternatives. The methodology produces surprisingly robust and attractive estimators. Further research would seem to be justified in order to deal with other specific model contexts, especially the linear dynamics, and to learn about the operational characteristics of the GRF methodology. Its firm Bayesian underpinning and the available evidence suggest very good prospects.

References

Barndorff-Nielsen, O.F., 1978, Information and exponential families in statistical theory (Wiley, New York, NY).

Fisher, F.M., 1961, On the cost of approximate specification in simultaneous equation estimation, Econometrica 29, 139–170.

Hendry, D.F. and J.-F. Richard, 1987, Recent developments in the theory of encompassing, CORE discussion paper no. 8722.

Liu, T.C., 1960, Under identification, structural estimation and forecasting, Econometrica 28, 855–865.

Maasoumi, E., 1978, A modified Stein-like estimator for the reduced form parameters of simultaneous equations, Econometrica 46, 695–703.

Maasoumi, E., 1986, Reduced form estimation and prediction from uncertain structural models: A generic approach, Journal of Econometrics 31, 3–29.

Maasoumi, E. and J.-H. Jeong, 1988, A comparison of GRF and other reduced form estimators in simultaneous equations models, Journal of Econometrics 37, 115–134.

Maasoumi, E. and P.C.B. Phillips, 1982, On the behaviour of inconsistent instrumental variables estimators, Journal of Econometrics 19, 183–201.

Morimune, K. and T. Sawa, 1980, Decision rules for the choice of structural equations, Journal of Econometrics 14, 329–347.

Richard, J.-F., 1980, Models with several regimes and changes in exogeneity, Review of Economic Studies 47, 1–20.

Sargan, J.D., 1988, The existence of the moments of the estimated reduced form coefficients, in: E. Maasoumi, ed., Contributions to econometrics: J.D. Sargan, Vol. II (Cambridge University Press, Cambridge).

Spanos, A., 1986, Statistical foundations of econometric modelling (Cambridge University Press, Cambridge).

White, H., 1982, Maximum likelihood estimation of misspecified models, Econometrica 50, 1–25.