Finite sample performance of the robust Wald test in simultaneous equation systems

Calzolari, Giorgio and Panattoni, Lorenzo

IBM Scientific Center, Pisa, Italy

1987

FINITE SAMPLE PERFORMANCE OF THE ROBUST WALD TEST IN SIMULTANEOUS EQUATION SYSTEMS

Giorgio Calzolari and Lorenzo Panattoni

The estimator of the coefficient covariance matrix proposed in White (1982) can be used to *robustify* the classical *Wald* test. Sampling experiments recently performed on linear regressions and simultaneous equation models, however, suggest that such an estimator *tends to underestimate* the covariance matrix if the model is correctly specified. In the classical framework of simultaneous equation systems, this chapter aims at investigating the consequences of the use of robust covariance matrix estimators in the *Wald* test when there is no misspecification.

I. INTRODUCTION

The robust estimator of the coefficient covariance matrix discussed in White (1980) for the linear regression model with heteroscedastic errors, then extended in White (1982) and Gourieroux et al. (1984) to cover more general types of models and misspecification, has become more and more popular in the last few years. It can easily be introduced into widely adopted computer programs, and its use in practical applications is often recommended. As a natural consequence, those tests that can use the robust covariance matrix estimator (*quasi-t*, *Wald*, and *Lagrange multiplier* tests) could be more appealing than tests that are usually *nonrobust* against misspecification (*likelihood ratio* and the traditional versions of the *Wald* and *Lagrange multiplier* tests).

However, several recent studies have indicated that the robust estimator *tends to underestimate* the coefficient covariance matrix if the model is correctly specified. Chesher and Jewitt (1984, 1987) identify conditions under which this covariance estimator is downwardly biased in the linear regression model (the *heteroscedasticity consistent* covariance estimator). They show that the bias critically depends on the regression design and can be severe. MacKinnon and White (1985) propose some finite sample corrections for this covariance estimator in linear regressions, whereas the sampling experiments in Prucha (1984) and in Calzolari and Panattoni (1984) clearly indicate a similar need for systems of simultaneous equations.

This chapter aims at investigating the small sample performance of the Wald test, based on the robust covariance estimator, when there is no misspecification.

We first briefly summarize explicit formulas for the likelihood and its first and second derivatives in a system of simultaneous equations with normal errors, then illustrate the covariance matrix estimator and its use in the Wald test briefly in Sections IV and V.

We then present detailed results of sampling experiments on a system of simultaneous equations taken from the literature (Klein-I model). Each group of experiments is performed on a sample period of different length and with different values of the exogenous variables (fixed across replications, but generated at the beginning with *platy-*, *meso-*, or *leptokurtic* distributions). For each Monte Carlo replication we compute the Wald test statistic for the hypothesis that *all* structural coefficients are equal to their *true* values. This involves reestimation of the structural model each time with *full information maximum likelihood* (*FIML*) and computation of the corresponding robust estimate of the coefficients covariance matrix.

We then draw several considerations from the experimental results. In Section VII, a scheme for explaining some behaviors of the robust Wald statistic is provided for the particular case of the linear regression model. We show that a poor performance of the test has to be expected when the explanatory (exogenous) variables exhibit, in the sample period, large moments of the fourth order. Experimental results on several other systems of simultaneous equations (Section VIII) confirm the important influence of the *sample kurtosis* of the exogenous variables on the small performance of the *robust Wald* test.

II. THE MODEL

We follow the notation of Amemiya (1977) for general nonlinear systems of simultaneous equations, with additive random error terms that are independently and identically distributed like multivariate normal. Refer to Amemiya's paper for details on the underlying assumptions. Let the simultaneous equation model be represented as

$$f_i(y_t, x_t, a_i) = u_{i,t} \quad i = 1, 2, \ldots, m; \; t = 1, 2, \ldots, T \tag{1}$$

where y_t is the $m \times 1$ vector of endogenous variables at time t, x_t is the vector of exogenous variables at time t, and a_i is the vector of unknown structural coefficients in the ith equation. The $m \times 1$ vector of random error terms at time t, $u_t = (u_{1,t}, u_{2,t}, \ldots, u_{m,t})'$, is assumed to be independently and identically distributed as $N(0, \Sigma)$ with Σ completely unknown, apart from being symmetric and positive definite. The complete $n \times 1$ vector of unknown structural coefficients of the system will be indicated as $a = (a_1', a_2', \ldots, a_m')'$.

III. THE LIKELIHOOD

The log-likelihood of the tth observation can be expressed as

$$L_t = -\tfrac{1}{2}\log|\Sigma| + \log\left\|\frac{\partial f_t}{\partial y_t'}\right\| - \tfrac{1}{2} f_t' \Sigma^{-1} f_t \tag{2}$$

where $f_t = (f_{1,t}, f_{2,t}, \ldots, f_{m,t})' = u_t$ and the Jacobian determinant $|\partial f_t / \partial y_t'|$ is taken in absolute value. The log-likelihood of the whole sample is

$$L_T = \sum_{t=1}^{T} L_t. \tag{3}$$

For the ith equation, we define $g_{i,t} = \partial f_{i,t}/\partial a_i$, which is a column vector with the same length as a_i; for any i and j, we also define the matrix $g_{i,j,t} = \partial^2 f_{i,t}/\partial a_i \partial a_j'$. If $i \neq j$, then $g_{i,j,t}$ is zero; it is zero also for $i = j$ if the model is linear in the coefficients (even if nonlinear in the variables). We note now that $g_{i,t}$ and $g_{i,j,t}$ may be regarded as functions of u_t, x_t, and a under the standard assumption of a one-to-one correspondence between u_t and y_t. Differentiating with respect to the coefficients of the ith equation, we get

$$\frac{\partial L_t}{\partial a_i} = \frac{\partial g_{i,t}}{\partial u_{i,t}} - g_{i,t} f_t' \sigma^i \tag{4}$$

where σ^i represents the ith column of Σ^{-1}. Differentiating with respect to the elements of Σ^{-1}, we get

$$\frac{\partial L_t}{\partial (\Sigma^{-1})} = \tfrac{1}{2}\Sigma - \tfrac{1}{2} f_t f_t' \tag{5}$$

where use has been made of $\partial g_{i,t}/\partial u_{i,t} = (\partial g_{i,t}/\partial y_t')(\partial f_t/\partial y_t')_i^{-1}$. No restriction has yet been placed on Σ; considering that Σ^{-1} is symmetric, differentiating with respect to its i,jth term we get

$$\frac{\partial L_t}{\partial \sigma^{i,j}} = \tfrac{1}{2}\sigma_{i,j} - \tfrac{1}{2} f_{i,t} f_{j,t} \quad (\times 2, \text{ if } i \neq j). \tag{6}$$

Further differentiation of (4) gives

$$\frac{\partial^2 L_t}{\partial a_i \partial a_j'} = \frac{\partial g_{i,i,t}}{\partial u_{i,t}} - \frac{\partial g_{i,t}}{\partial u_{j,t}} \frac{\partial g_{i,t}'}{\partial u_{i,t}} - g_{i,i,t} f_t' \sigma^i - \sigma^{i,j} g_{i,t} g_{j,t}' \tag{7}$$

$$\frac{\partial^2 L_t}{\partial \sigma^{i,i} \partial a_i} = -g_{i,t} f_{i,t} \tag{8}$$

$$\frac{\partial^2 L_t}{\partial \sigma^{i,j} \partial a_i} = -g_{i,t} f_{j,t} \tag{9}$$

$$\frac{\partial^2 L_t}{\partial \sigma^{i,j} \partial a_j} = -g_{j,t} f_{i,t} \tag{10}$$

$$\frac{\partial^2 L_t}{\partial \sigma^{i,j} \partial a_r} = 0 \quad \text{if } r \neq i \text{ and } r \neq j \tag{11}$$

$$\frac{\partial^2 L_t}{\partial \sigma^{i,j} \partial \sigma^{r,r}} = -\tfrac{1}{2}\sigma_{i,r}\sigma_{r,j} \quad (\times 2, \text{ if } i \neq j) \tag{12}$$

$$\frac{\partial^2 L_t}{\partial \sigma^{i,j} \partial \sigma^{r,s}} = -\tfrac{1}{2}\sigma_{i,r}\sigma_{s,j} - \tfrac{1}{2}\sigma_{i,s}\sigma_{r,j} \quad \text{if } r \neq s \ (\times 2, \text{ if } i \neq j). \tag{13}$$

For models that are linear in the coefficients (even if nonlinear in the variables), $g_{i,i,t}$ and its derivatives are zero, so that the first and third term on the right-hand side of Eq. (7) vanish. Moreover, $-g_{i,t}$ is nothing but the vector of values, at time t, on the explanatory variables of the ith equation. Therefore, the numerical evaluation of all the above equations requires only one order of differentiation: the computation of derivatives of the explanatory endogenous variables in the ith and jth equations with respect to the error terms of the same equations. Furthermore, since $\partial g_{i,t}/\partial u_{i,t} = (\partial g_{i,t}/\partial y_t')(\partial f_t/\partial y_t')_i^{-1}$, this differentiation could even be performed analytically without any particular difficulty. The use of Eqs. (7–13) for the computation of the Hessian matrix is, therefore, a sufficiently manageable matter even for medium-large models.

The formulas given above can be used to build all the matrices used in this study.

IV. ESTIMATORS OF THE COVARIANCE MATRIX

Using the formulas of the previous section, we can build several estimators of the asymptotic covariance matrix. We shall first deal with the vector of all unknown structural parameters. We may stack the estimated coefficients $\hat{a}$ and the elements of the estimated $\hat{\Sigma}^{-1}$ into a column vector of estimated parameters, $\hat{p}$. Obviously, since $\hat{\Sigma}^{-1}$ is symmetric, we shall stack into this vector only the columns of a triangular part of $\hat{\Sigma}^{-1}$ (operator *vech*)

$$\hat{p} = \begin{bmatrix} \hat{a} \\ vech\,\hat{\Sigma}^{-1} \end{bmatrix}. \tag{14}$$

In this way, with n being the number of unknown structural coefficients and m the number of stochastic equations, the length of the whole vector of parameters p is $n + m(m+1)/2$; the whole information matrix (and the asymptotic covariance matrix of $\hat{p}$) has dimensions $[n + m(m+1)/2] \times [n + m(m+1)/2]$.

Equations (7–13), with the minus sign and summed over the sample period, provide the elements or the blocks of the $[n + m(m+1)/2] \times$

$[n + m(m+1)/2]$ Hessian matrix of the log-likelihood

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} = -\sum_{t=1}^{T} \frac{\partial^2 L_t}{\partial p \partial p'} = -\frac{\partial^2 L_T}{\partial p \partial p'}$$

$$= -\begin{bmatrix} \frac{\partial^2 L_T}{\partial a \partial a'} & \frac{\partial^2 L_T}{\partial a \partial (vech\,\Sigma^{-1})'} \\ \frac{\partial^2 L_T}{\partial (vech\,\Sigma^{-1}) \partial a'} & \frac{\partial^2 L_T}{\partial (vech\,\Sigma^{-1}) \partial (vech\,\Sigma^{-1})'} \end{bmatrix}. \quad (15)$$

Equations (4–6) provide the first derivatives of the log-likelihoods with respect to all the unknown structural form parameters. We may get an estimate of the whole information matrix by computing the outer product of the first derivatives

$$B = \begin{bmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{bmatrix} = \sum_{t=1}^{T} \frac{\partial L_t}{\partial p} \frac{\partial L_t}{\partial p'}$$

$$= \sum_{t=1}^{T} \begin{bmatrix} \frac{\partial L_t}{\partial a} \frac{\partial L_t}{\partial a'} & \frac{\partial L_t}{\partial a} \frac{\partial L_t}{\partial (vech\,\Sigma^{-1})'} \\ \frac{\partial L_t}{\partial (vech\,\Sigma^{-1})} \frac{\partial L_t}{\partial a'} & \frac{\partial L_t}{\partial (vech\,\Sigma^{-1})} \frac{\partial L_t}{\partial (vech\,\Sigma^{-1})'} \end{bmatrix}. \quad (16)$$

When all derivatives in (15) and (16) are calculated in $\hat{p}$, the resulting matrices will be indicated as $\hat{A}$ and $\hat{B}$. If $\hat{p}$ is the *full information maximum likelihood* (*FIML*) estimate of the parameters vector, it is well known (see, for example, Rothenberg, 1973, pp. 10–11) that, under correct specification of the model and of the error generating process, $\hat{A}/T$ and $\hat{B}/T$ converge asymptotically to the information matrix, and their inverse converge to the asymptotic covariance matrix of the structural parameters. Therefore, both of them can be used for constructing tests that will have the right size with large samples.

When the distribution of the random error process does not coincide with the distribution underlying the likelihood, FIML estimation may nevertheless provide estimates of the parameters that are still consistent, but the two traditional expressions for the information matrix are no longer equivalent, and their inverses generally provide inconsistent estimates of the parameters covariance matrix. Under assumptions with a varying degree of generality, White (1982, 1983, for independently and identically distributed observations), Gourieroux et al. (1984, for independently, nonidentically distributed observations), and Domowitz and White (1982, for dependent and heterogeneously distributed observations) derive the estimator of the covariance matrix

$$\hat{C} = \begin{bmatrix} \hat{C}_{1,1} & \hat{C}_{1,2} \\ \hat{C}_{2,1} & \hat{C}_{2,2} \end{bmatrix} = \hat{A}^{-1}\hat{B}\hat{A}^{-1}$$

$$= \left[-\left.\frac{\partial^2 L_T}{\partial p \partial p'}\right|_{\hat{p}} \right]^{-1} \left[\sum_{t=1}^{T} \left.\frac{\partial L_t}{\partial p} \frac{\partial L_t}{\partial p'}\right|_{\hat{p}} \right] \left[-\left.\frac{\partial^2 L_T}{\partial p \partial p'}\right|_{\hat{p}} \right]^{-1} \quad (17)$$

which is generally consistent regardless of whether or not the actual distribution of the error process coincides with that underlying the likelihood. This matrix extends to more general classes of models the *heteroskedasticity consistent* estimator of the covariance matrix proposed in White (1980) for the linear regression model (we shall come again to this point in Section VII). Tests that make use of this *robust* estimator of the covariance matrix will have the right size for large samples.

If the error process coincides with that underlying the likelihood, as supposed in this chapter, $T\hat{C}$ will also asymptotically converge to the inverse of the information matrix as well as $T\hat{A}^{-1}$ and $T\hat{B}^{-1}$.

V. DESIGN OF THE MONTE CARLO EXPERIMENTS

In all the experiments described in this chapter we examine the small sample performance of the *robust Wald* (*RW*) test when the hypothesis being tested is that structural coefficients assume given values. All sampling experiments are performed starting from a "*true*" vector of coefficients and calculating the test statistic that *all* coefficients are equal to their *true* values.

Experiments have been performed on several small to medium sized models. The models, taken from the literature, maintain the structure of real world models. For each model, given the set of *true* parameters (coefficients and covariance matrix of the structural disturbances, held fixed over all the replications), we fix a sample period length and generate values of the exogenous variables over the sample period: platy-, meso-, and leptokurtic distributions are used in different experiments. Whichever generation method has been used, the sample of exogenous variables is then kept fixed in all the Monte Carlo replications of each experiment. Also experimented with are the historical *real-world* values of the exogenous variables taken from the literature and repeated consecutively for the long sample cases.

Each Monte Carlo replication proceeds as follows. Independently of the exogenous variables we generate random values of the structural disturbances over the sample period. Obviously, the distribution for this random error process *must* be multivariate normal with zero mean and the given covariance matrix. Finally we compute the values of the endogenous variables with stochastic simulation over the sample period.

We now compute FIML estimates[1] of the structural parameters $\hat{a}$ and $\hat{\Sigma}$ and the robust covariance matrix estimate $\hat{C}$ as in (17). In general, if we wish to test the hypothesis H_0: $s(p) = 0$, where the restrictions are formulated as a vector function of the structural parameters, and $\nabla s(p)$ is the Jacobian of the vector of restrictions, the appropriate form of the robust Wald test statistic, given in theorem 3.4 in White (1982), is

$$RW = s(\hat{p})' [\nabla s(\hat{p}) C(\hat{p}) \nabla s(\hat{p})']^{-1} s(\hat{p}). \quad (18)$$

Since the particular restriction being tested here is that the vector of all structural coefficients is equal to a, we have $s(\hat{p}) = (\hat{a} - a)$. The Jacobian $\nabla s(\hat{p})$ assumes, therefore, the form of an $n \times n$ unit matrix followed by a matrix of zeroes, and the robust Wald test statistic simply becomes

$$RW = (\hat{a} - a)' (\hat{C}_{1,1})^{-1} (\hat{a} - a) \quad (19)$$

where $\hat{C}_{1,1}$ is the first $n \times n$ block of C, computed at the point that maximizes the likelihood.

Using the Hessian estimator of the covariance matrix with (19), we get the *Hessian Wald* (*HW*) test statistic

$$HW = (\hat{a} - a)' (\hat{A}^{1,1})^{-1} (\hat{a} - a) \quad (20)$$

where $\hat{A}^{1,1}$ is the first block of the inverse of matrix $\hat{A}$ [Eq. (15); see also Engle, 1984, Eq. (11)][2].

For the *likelihood ratio* (*LR*) test, since the null hypothesis restricts the value of all the structural form coefficients, a, *constrained FIML* estimation is confined to the computation of the σ parameters. Computation of the parameters that maximize the *constrained* likelihood is performed by simply plugging the "*true*" coefficients (under H_0) into the model, then computing the corresponding structural residuals; finally the σ parameters are computed from the usual cross-products of these residuals. The values of the log-likelihood function, computed at the *unconstrained* and *constrained* maximum points, give the likelihood ratio test statistic

$$LR = 2[L_T(\hat{a}) - L_T(a)]. \quad (21)$$

For each model, for each sample period length, and for each different generation process of the exogenous variables we perform a few hundred replications of the Monte Carlo process obtaining the small sample distribution of the HW, the RW, and of the LR test statistics.

The three tests are asymptotically equivalent if the model is correctly specified. Otherwise, the likelihood ratio and the Hessian Wald tests do not have the correct asymptotic size and generally fail to converge to a χ^2 distribution (White, 1982). In other words, they are generally *nonrobust* against misspecification of the random error process. In sampling experiments, in a correct specification framework as in this chapter the behavior of the LR and the HW statistics can be used for comparison with the RW test in small samples. Their better fit with the asymptotic χ^2 distribution, which will be evidenced in all our simulation results, clearly shows that asymptotic robustness in the Wald test has a cost in terms of finite sample performance.

VI. A CASE STUDY: KLEIN-I MODEL

In this section we describe in some detail the results of the experiments performed on a system of simultaneous equations whose structure is that of Klein's model-I (Klein, 1950). The *qualitative* behavior of the results, however, is not changed very much by changing the model, as will be clear from the summary tables of Section VIII. The structural form of the model is the following:

$$\begin{aligned} C_t &= a_1 + a_2 P_t + a_3 P_{t-1} + a_4 (W1 + W2)_t + u_{1,t} \\ I_t &= a_5 + a_6 P_t + a_7 P_{t-1} + a_8 K_{t-1} + u_{2,t} \\ W1_t &= a_9 + a_{10}(Y + T - W2)_t \\ &\quad + a_{11}(Y + T - W2)_{t-1} + a_{12} t + u_{3,t} \\ Y_t &= C_t + I_t + G_t - T_t \\ P_t &= Y_t - W1_t - W2_t \\ K_t &= K_{t-1} + I_t \end{aligned} \quad (22)$$

Number of equations = 6.
Number of stochastic equations $m = 3$.
Number of structural unknown coefficients $n = 12$.
Number of structural unknown parameters $n + m(m + 1)/2 = 18$.

As "*true*" values of the unknown parameters, we use the two stage least-squares estimates based on the 21-year sample period, 1921–1941, in Rothenberg (1973, Chap. 5).

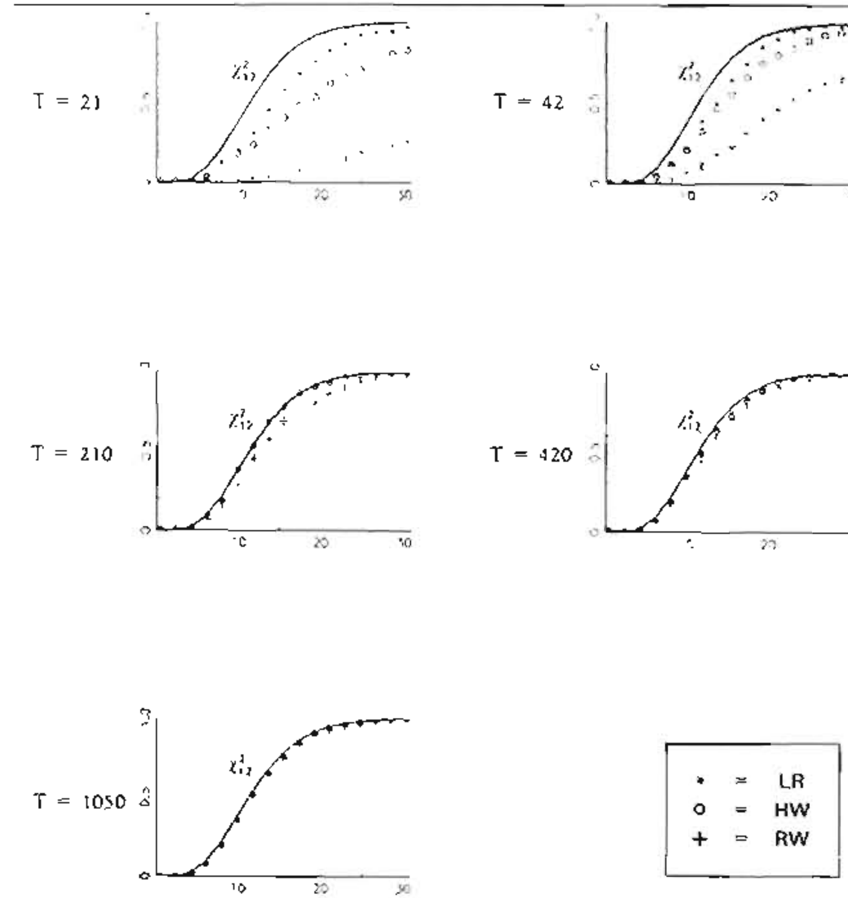
The model is dynamic; however, this should not raise particular problems, since we are operating in a correct specification framework (see White, 1983). In any case, all experiments are repeated twice. In one case, the model is treated as static in which variables that appear as lagged endogenous are replaced by current exogenous variables. In the other case, the model is treated as dynamic; therefore in each replication, the model is dynamically solved over the sample period, and the simulated values of the lagged endogenous variables are used when reestimating the structural parameters. Again, it will be clear that the *quality* of the results does not change. Convergence to the appropriate asymptotic distribution will be evidenced in almost all experiments, although, in some cases, this will require very long samples.

In the first group of experiments, we use the 21 historical observations of the exogenous variables (and of the lagged endogenous for the static case). Larger samples lengths (42, 63, etc.) are obtained replicating the same 21 observations.

The figures in Tables 1 and 2 display the experimental results in terms of cumulated distribution functions. The continuous curve corresponds to the asymptotic distribution of the test statistics, that is $\chi^2_{(12)}$. Each figure is related to 500 Monte Carlo replications, and the curves are smoothed by joining 17 points of each distribution. Some experiments have also been performed with a larger number of replications—up to 10 000. Since none of them evidenced any substantial difference from the overall behavior of the small sample distributions, the slight gain in accuracy did not seem to compensate the much higher cost of the experiment.[3]

The LR test does not perform too badly in the short sample, at least as far as the entire distribution is concerned. A more careful inspection of the critical region (the rightmost tail of each curve) shows that both tests give a probability of type-I errors slightly[4] larger than the nominal (asymptotic) size of the tests. For example, for $T = 42$, and nominal sizes 10 and 5%, the estimated rejection probabilities are

Table 1. Klein's Model-I. Static—Historical Exogenous Variables

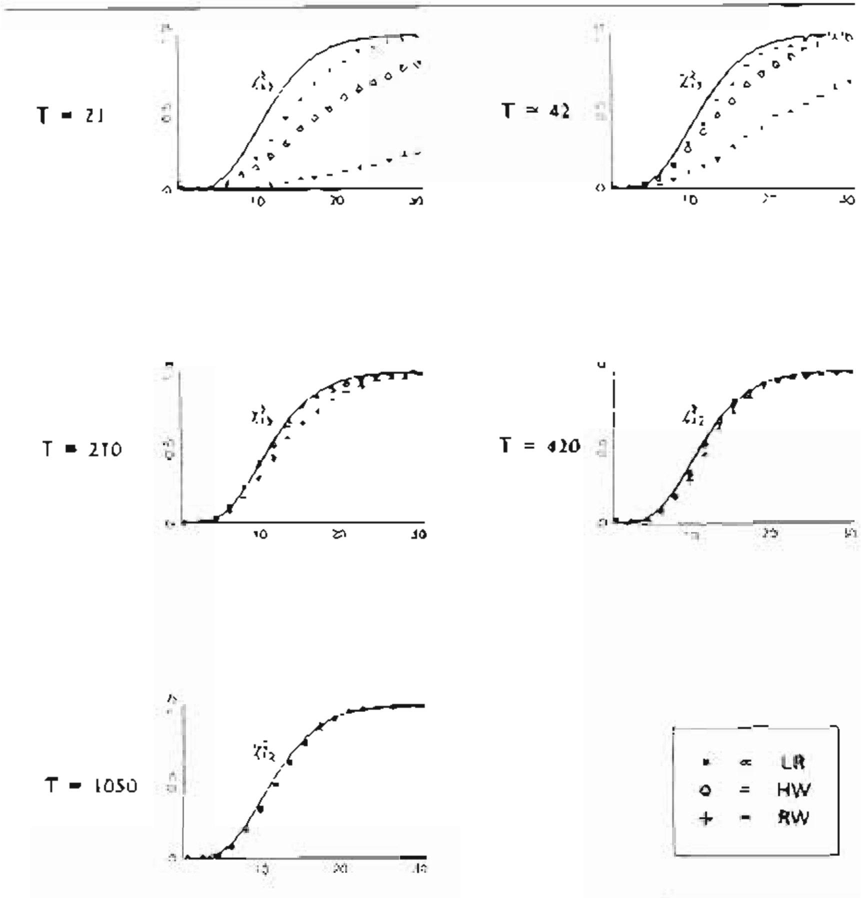


[a]Small sample distribution of LR, RW and HW test statistics $\chi^2_{(12)}$.

approximately 19 and 9% for the likelihood ratio test. The performances of the HW test are less brilliant: 25 and 17% are the expected probabilities of type-I errors corresponding to nominal sizes 10 and 5%. The discrepancy is larger for the historical period, but $T = 21$ is presumably too short for a model with 12 coefficients. The performance of the RW test is even less brilliant. Of course, it improves as the sample is enlarged. A sample period with 420 observations makes all small sample distributions hardly distinguishable from the asymptotic ones.

An important observation, as will be clear in the next section, is related to the *kurtosis* of the exogenous variables in the sample period

Table 2 Klein's Model-I: Dynamic—Historical Exogenous Variables



*Small sample distribution of LR, RW and HW test statistics $\chi^2_{11.5}$

(more precisely, the ratio between the fourth moment and the squared second moment about zero; it would be the kurtosis if the variables were previously normalized); its value varies between 1 (for the constant) and 1 8 (for the variable G)

It is interesting to observe that the corresponding static and dynamic cases do not exhibit substantial differences of behavior.

A characteristic common to all the cases is the relative position of the distributions. The right-most sampling distribution is that of the RW test statistic; the left-most is that of the LR, and the distribution of HW test statistic is between the other two. All three are right shifted with respect to the asymptotic χ^2.

The three small sample distributions are right shifted from the χ^2, thus implying for each test a probability of type-I errors greater than the nominal size of the test. Perhaps it is not clear enough from the results what happens in the very right-most part of the critical region, and, therefore, an analysis of the tests' behavior at 1% would require sampling experiments with more replications and the use of some suitable computational method to reduce the sampling variability, such as Davidson and MacKinnon's (1981) control variates. At 10 or 5%, however, there seem to be no doubts that the expected rejection probability is greatest for the RW test, less for HW, and least for the LR test.

The fact that the expected rejection probability is larger than the nominal size for the Wald test is consistent with results previously obtained by Calzolari and Panattoni (1984, 1988) on systems of simultaneous equations and with the results obtained by MacKinnon and White (1985) on a linear regression *without correcting factors* on the robust covariance matrix estimator.

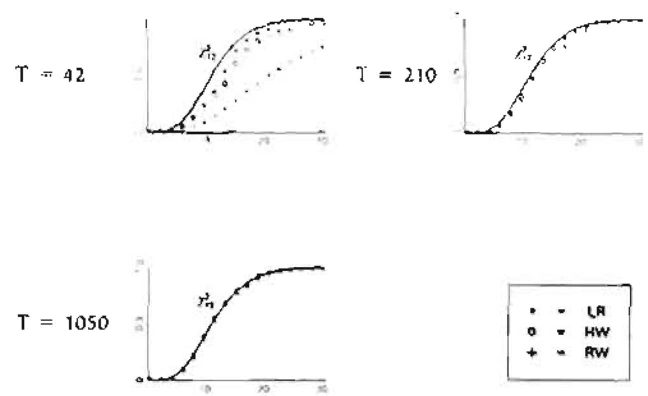
Another interesting consideration that can be drawn from Tables 1 and 2 is that the relative position of the sampling distributions of the likelihood ratio and the Wald statistics is consistent with the traditional inequality LR $\leq$ W derived by Savin (1976), Berndt and Savin (1977), and Evans and Savin (1982) for the linear regression and multivariate linear regression models. This suggests that the inequality holds *on average* even for simultaneous equations systems (where it does not hold algebraically in each replication given the nonzero correlation between estimated coefficients and σ parameters, as observed in Breusch, 1979).

We now perform a second group of experiments without using the historical observations of the exogenous variables but with randomly generated values. However, as already indicated, the scheme of the experiment is still with fixed exogenous variables, since they are generated once at the beginning and then kept fixed in all replications We first adopt a multivariate normal distribution with means and covariance matrix taken from the historical sample

$$\tilde{x}_t = T^{-1} \sum_{t=1}^{T} x_t e_t + \bar{x} \qquad (23)$$

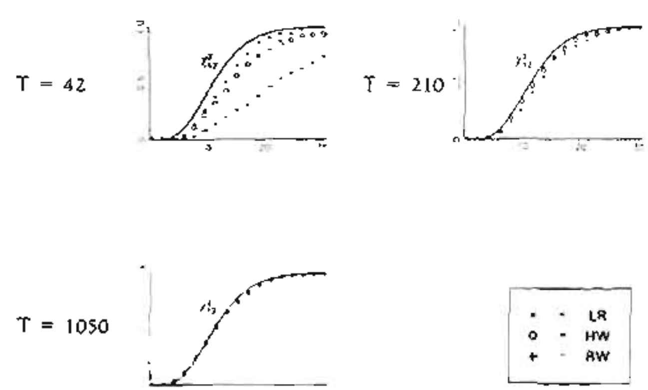
where the scalar random numbers e_t are iid $N(0, 1)$. Normality implies a *sample kurtosis* for each generated exogenous approximately = 3. A sample of exogenous variables with length T is generated at the

Table 3. Klein's Model-I: Static—Normal Exogenous Variables



[a]Small sample distribution of LR, RW and HW test statistics. $\chi^2_{(12)}$.

Table 4. Klein's Model-I: Dynamic—Normal Exogenous Variables



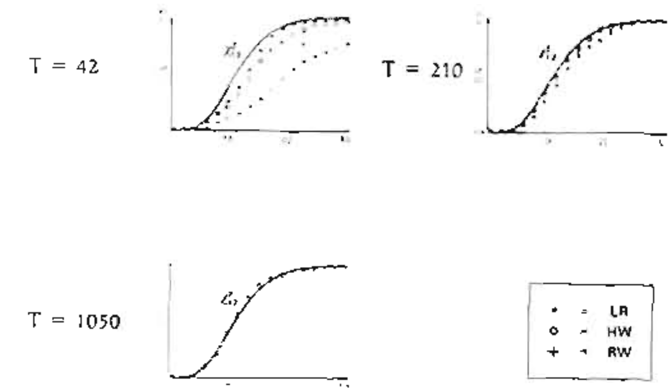
[a]Small sample distribution of LR, RW and HW test statistics. $\chi^2_{(12)}$.

beginning and then kept fixed over the 500 replications of each experiment.

Table 3 displays the results related to sample period lengths $T = 42$, 210, and 1050 for the static version of the model. Table 4 displays corresponding results for the dynamic model There are no remarkable differences between results from the two models.

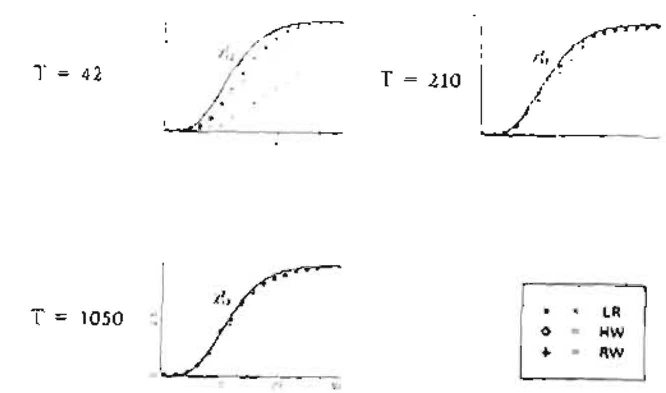
A new group of experiments is performed by generating random values of the exogenous variables with platykurtic distribution (*kurtosis* < 3). This is obtained with a simple modification of the

Table 5. Klein's Model-I: Static—Platykurtic Exogenous Variables



[a]Small sample distribution of LR, RW and HW test statistics. $\chi^2_{(12)}$.

Table 6. Klein's Model-I: Dynamic—Platykurtic Exogenous Variables



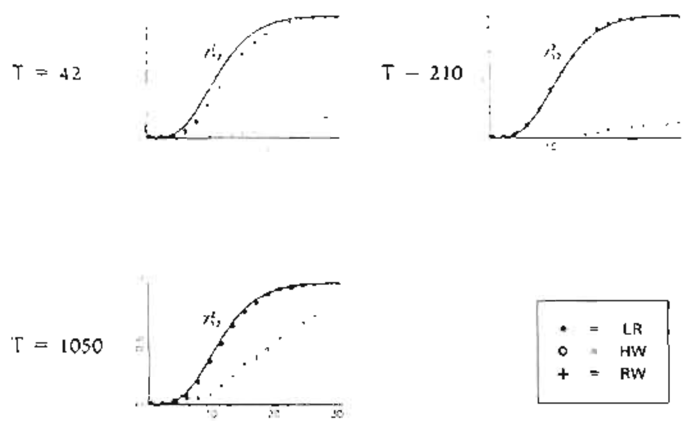
[a]Small sample distribution of LR, RW and HW test statistics. $\chi^2_{(12)}$.

generator (23)

$$\tilde{x}_t = T^{-1} \sum_{t=1}^{T} x_t |e_t|^c \operatorname{sign}(e_t) + \bar{x} \tag{24}$$

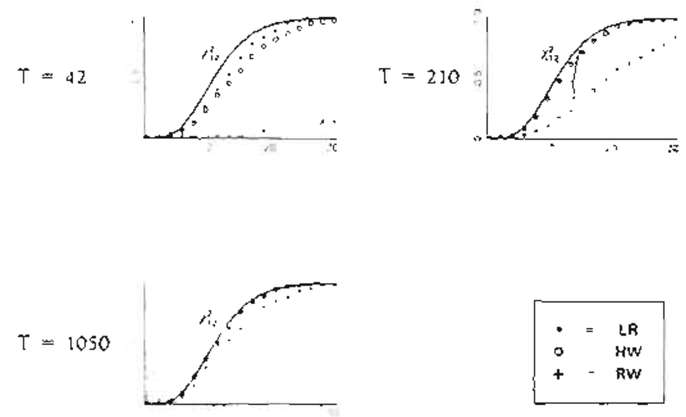
still with e_t's generated iid $N(0, 1)$ and with a value of the exponent c less than 1 (0.2 in our experiments; this gave values of the kurtosis less than 1.5 for all variables). The results displayed in Tables 5 and 6 show a very slight improvement in the behavior of the RW test statistic, whose distribution approximates faster the asymptotic one.

Table 7. Klein's Model-I: Static—Leptokurtic Exogenous Variables



aSmall sample distribution of LR, RW and HW test statistics, $\chi^2_{(12)}$.

Table 8. Klein's Model-I: Dynamic—Leptokurtic Exogenous Variables



aSmall sample distribution of LR, RW and HW test statistics, $\chi^2_{(12)}$.

The last group of experiments, displayed in Tables 7 and 8, has been performed generating random values of the exogenous variables with a distribution strongly leptokurtic. The generator adopted is

$$\tilde{x}_t = T^{-1} \sum_{t=1}^{T} x_t e_t^c + \bar{x} \tag{25}$$

with an odd value of $c > 1$. We have used a rather large value of $c = 7$ for better exemplifying the behavior of the test statistics. This has produced samples of exogenous variable with *kurtosis* up to 16. This seems to have no particular effect on thedistributions of LR and HW test statistics, but has a dramatic impacton the RW statistic. For the static model with a rather long sample ($T = 210$, 10 times longer than the historical sample), for a nominal siz of 5%, the estimated rejection probability is over 90%. Still with avery long sample ($T = 1050$, 50 times longer than the historical sanple) at 5%, the estimated rejection probability is almost 50%. Tle discrepancy in this case is considerably larger for the static moel (Table 7), where all the predetermined variables, including lagd endogenous, are strongly leptokurtic, than for the dynamic mode (Table 8).

VII. THE SAMPLE KURTOSIS OF EXOGENOUS VARIABLES

We start from consideration of a linea regression model with one explanatory variable

$$y_t = a x_t + u_t; \quad u_t \text{ iid} N(0, \sigma^2). \tag{26}$$

With the notations of Sections III and IV, the vector of unknown parameters is

$$p = [a, \sigma^{-2}]' \tag{27}$$

while $g_{t,t} = -x_t$. FIML estimators are

$$\hat{a} = \left(\sum_{t=1}^{T} x_t^2\right)^{-1} \sum_{t=1}^{T} x_t y_t; \quad \hat{\sigma}^{-2} = T\left(\sum_{t=1}^{T} \hat{u}_t^2\right)^{-1}. \tag{28}$$

As is well known, the Hessian of the loglikelihood computed at the point which maximizes the likelihood is diagonal ($\hat{A}_{1,2} = 0$ and $\hat{A}_{2,1} = 0$). From this it follows that

$$\hat{C}_{1,1} = \hat{A}^{1,1} \hat{B}_{1,1} \hat{A}^{1,1} = \left(\sum_{t=1}^{T} x_t^2\right)^{-2} \sum_{t=1}^{T} x_t^2 \hat{u}_t^2. \tag{29}$$

We have now

$$E(\hat{u}_t^2) = \sigma^2 \left[1 - \frac{x_t^2}{\sum_{t=1}^{T} x_t^2}\right]. \tag{30}$$

Therefore, computing the expected valu of (29), we get

$$E(\hat{C}_{1,1}) = \sigma^2 \left(\sum_{t=1}^{T} x_t^2\right)^{-1} \left[1 - \frac{\sum_{t=1}^{T} x_t^4}{(\sum_{t=1}^{T} x_t^2)^2}\right]$$
$$= \sigma^2 \left(\sum_{t=1}^{T} x_t^2\right)^{-1} \left(1 - \frac{k_x}{T}\right) \tag{31}$$

where k_x is the average fourth power of the explanatory variable divided by the square of its mean square. If the exlanatory variable has zero mean, k_x is the *sample kurtosis* of the explanatory variable. T being the length of the sample period, k_x may assume values between 1 and T. Under the assumptions given in (26), the variance of $(\hat{a} - a)$ is $\sigma^2 (\sum x_t^2)^{-1}$, and we easily see from Eq. (31) that $\hat{C}_{1,1}$ is in any case biased downward.

The amount of this bias in a practical application may be small enough to be negligible if k_x is small. For example, the smallest would be for a constant series where $k_x = 1$ yields the least bias, with a bias factor $1 - 1/T$, similar to that obtained by the traditional OLS variance estimator without degrees of freedom correction.

A platykurtic behavior of the explanatory variable $k_x < 3$ also will cause a downward bias, usually negligible, and the same can be said for a mesokurtic behavior ($k_x = 3$, as for x_t's generated by a normal process). Of course, whether bias is negligible or not depends on how large the sample is. For example, $T = 10$ may already be considered *sufficiently large* if there is only one coefficient to be estimated as in this example. In that case the variance would be biased downward by 30%, which may not be considered a small amount.

Obviously, the most dramatic effects are obtained when the explanatory variable is strongly *leptokurtic* ($k_x > 3$) as is clear from Eq. (31). Platykurtic explanatory variables are certainly more likely to be encountered in practical applications (e.g., constants, or variables that exhibit a constant growth over time), but large values of the kurtosis may also be encountered in some cases (e.g., seasonal *dummy* variables)[5].

Most of the argument above can be extended to the case of general linear regression models with more than one explanatory variable. In this case, if there is no misspecfication, it is no more true that the *robust* covariance matrix estimator is always more biased than the traditional estimator; we may have cases in which it is less biased. Bias, however, is always downward and it is larger when the fourth-order sample moments, or cumulants, of the explanatory variables are large. This will be clear if, for the linear regression with K explanatory variables

$$y = Xa + u; \quad X' = [x_1, x_2, \ldots, x_t, \ldots, x_T]; \quad u \sim N(0, \sigma^2 I) \tag{32}$$

we compute the FIML estimator (=OLS) of the coefficients, $\hat{a} = (X'X)^{-1}X'y$, then their robust covariance matrix estimator (recalling that the Hessian is still block diagonal),

$$\hat{C}_{1,1} = \hat{A}^{1,1}\hat{B}_{1,1}\hat{A}^{1,1} = (X'X)^{-1}\sum_{t=1}^{T} x_t x_t' \hat{u}_t^2 (X'X)^{-1} \tag{33}$$

which is equal to the *heteroskedasticity consistent* estimator given in White (1980). Recalling that

$$\hat{u} = [I - X(X'X)^{-1}X']u \tag{34}$$

and computing the expected value of (33) we get

$$E(\hat{C}_{1,1}) = \sigma^2(X'X)^{-1} - \sigma^2(X'X)^{-1}\sum_{t=1}^{T}[x_t x_t'(X'X)^{-1}x_t x_t'](X'X)^{-1}. \tag{35}$$

It is clear that the estimator is biased downward, since the covariance matrix of $\hat{a} - a$ is equal to the first term on the right-hand side of Eq. (35); however, the second term is not necessarily greater than or equal to $\sigma^2(X'X)^{-1}K/T$ in matrix sense [we can only prove that it is always $\geq \sigma^2(X'X)^{-1}/T$]. Thus, the bias is not necessarily greater than the one given by the Hessian covariance matrix estimator, that is, the traditional OLS formula, without degrees of freedom correction.

Equation (35) also makes clear the role of the fourth powers of the x_t's: the higher they are, the larger the second term on the right-hand side of (35), and the smaller the expectation of the *robust* covariance matrix estimator.

The robust covariance matrix estimator in (33) must be inverted to compute the RW test statistic on a vector of coefficients [see Eq. (19)]. Cases in which the estimator shows strong downward bias are likely to produce absolute values *too large* of the statistic. Therefore, for the RW test statistic we should expect, in these cases, a right shift of the small sample distribution with respect to its asymptotic distribution (χ^2). The probability of type-I error will be larger than the nominal size of the test.

For a system of simultaneous equations, we are unable to give a simple interpretation of the phenomenon analytically. The sampling experiments described in the previous and following sections suggest that something similar to the linear regression model is likely to occur also in that case. Exogenous variables with a large fourth-order moment cause a strong downward bias to the robust covariance matrix estimator, and, therefore, imply values of the RW test statistic that tend to be too large.

This also follows from the Monte Carlo experiments described in Prucha (1984) and in Calzolari and Panattoni (1984, 1988). It was shown there that the covariance matrix estimator based on the outer product of the first-order derivatives of the likelihood ($\hat{B}^{-1}$) tends to be systematically larger than the Hessian estimator $\hat{A}^{-1}$. Therefore, $\hat{A}^{-1}\hat{B}$ tends to be smaller than the unit matrix, and $\hat{C}$ tends to be even smaller than the Hessian covariance estimator. This implies a value of the RW statistic usually larger than the corresponding value of the HW statistic.

Finally we observe that in the linear regression model a large fourth-order moment of the exogenous variables strongly affects the estimator $\hat{B}$, while it has no effect on $\hat{A}$. The sampling experiments performed for the present study (not included in the chapter for brevity's sake) confirm that the influence of leptokurtic exogenous variables is on the outer product covariance estimator, and that their influence on the robust covariance matrix estimator follows as a consequence.

We may conclude that asymptotic robustness for the Wald test may have a large cost in terms of finite sample performance, mainly depending on the behavior of the exogenous variables. In models that involve strongly leptokurtic variables (for example, seasonal *dummy* variables), the use of the robust covariance matrix estimator in Wald test does not seem to be recommended. Correcting factors are needed for simultaneous equation systems, analogous to those proposed by MacKinnon and White (1985) for the covariance matrix in the linear regression model.

VIII. RESULTS ON OTHER MODELS

In this section we briefly summarize the results of experiments performed on three other simultaneous equation systems. Two are small models of the Italian economy proposed in the literature, and one is a nonlinear version of Klein's model-I. The following remarks apply to all models and experiments.

1. Models have been treated either as static models (with lagged endogenous variables replaced by fixed current exogenous) or as dynamic ones. No particular divergence has been shown by the experiments, including the case of the nonlinear dynamic model. For brevity's sake, the tables of results refer only to the static version of the models.
2. As with the experiments discussed in Section VI, we display the sampling distributions of the *likelihood ratio* test statistic and of the *robust* and *Hessian* versions of the *Wald* test statistic.
3. For each model in static version, we repeat the experiment 12 times, that is, with 4 different sets of exogenous variables (historical—repeated consecutively for long samples, normal, platykurtic, and leptokurtic) and 3 different lengths of the sample period.

The results, which are displayed in Tables 9–20, suggest considerations that are essentially the same as those derived for Klein's model-I in Section VI. The only behavior remarkably different is in the Sitzia and Tivegna (1975) model for the Italian economy using the historical values of the exogenous variables. The presence in the model of a *dummy* variable (all values are zero, except in 3 years; therefore, k_x is much greater than 3) seems sufficient to explain this behavior in light of the considerations of Section VII.

Table 9. A Simple Macroeconomic Model of the Italian Economy[a]

$$C_t = a_1 + a_2 Y_t + a_3 C_{t-1} + u_{1,t}$$
$$I_t = a_4 + a_5(Y_t - Y_{t-1}) + a_6 I_{t-1} + u_{2,t}$$
$$M_t = a_7 + a_8 I_t + a_9(Y - I)_t + u_{3,t}$$
$$Y_t = C_t + I_t + Z_t - M_t$$

Number of equations = 4
Number of stochastic equations $m = 3$
Number of structural unknown coefficients $n = 9$
Number of structural unknown parameters $n + m(m+1)/2 = 15$

[a] The model, specifically designed for the purpose of analyzing the effects of current revisions in Italian national account series, is described in Rettore and Trivellato (1986).

Table 10. Simple Italian Model ($\chi^2_{(9)}$; $T = 19$)

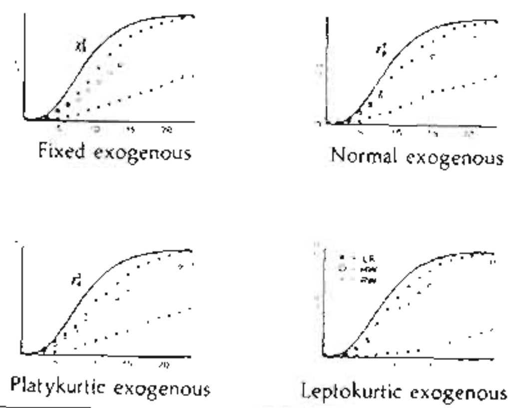
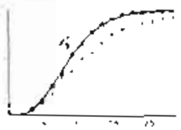
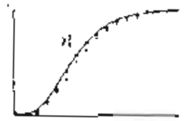


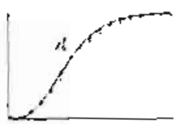
Table 11. Simple Italian Model ($\chi^2_{(9)}$; $T = 200$)



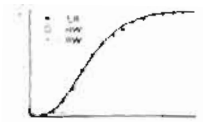
Fixed exogenous



Normal exogenous



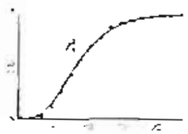
Platykurtic exogenous



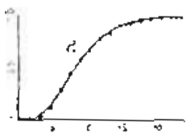
Leptokurtic exogenous

Table 12. Simple Italian Model ($\chi^2_{(9)}$, $T = 1000$)

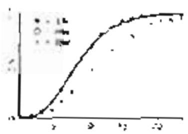
Fixed exogenous



Normal exogenous



Platykurtic exogenous



Leptokurtic exogenous

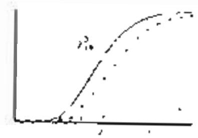
Table 13. A Linear Model for the Italian Economy[a]

$$\begin{aligned}
CPN_t &= a_1 + a_2(WIT + WG + X2)_t \\
&\quad + a_3(PIT + PAF)_t + a_4(PIT + PAF)_{t-1} + u_{1t}, \\
ILIT_t &= a_5 + a_6 PIT_{t-1} + a_7 KOCC_t + a_8 ILIT_{t-1} + u_{2t}, \\
M_t &= a_9 + a_{10}(CPN + ILIT)_t + a_{11} t + u_{3t}, \\
WIT_t &= a_{12} + a_{13}(WIT + PIT)_t + a_{14} KOCC_t \\
&\quad + a_{15} DUS70_t + a_{16} WIT_{t-1} + u_{4t}, \\
KOCC_t &= a_{17} + a_{18}(ILIT_t + ILIT_{t-1} + ILIT_{t-2}) \\
&\quad + a_{19}(ILIT_{t-1} + 2 \times ILIT_{t-2}) + u_{5t}, \\
PIT_t &= RNLCF_t - WIT_t - WG_t - PAF_t - X2_t, \\
RNLCF_t &= CPN_t + ILIT_t - M_t + WG_t - TI_t + X1_t
\end{aligned}$$

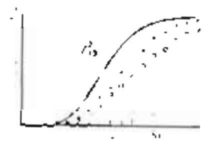
Number of equations = 7
Number of stochastic equations $m = 5$
Number of structural unknown coefficients $n = 19$
Number of structural unknown parameters $n + m(m + 1)/2 = 34$

[a] Model, meaning of the variables, and data for the Italian economy 1952–1971 can be found in Sitzia and Tivegna (1975)

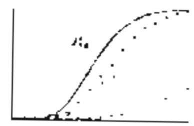
Table 14. Linear Italian Model ($\chi^2_{(19)}$; $T = 40$)



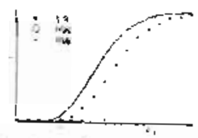
Fixed exogenous



Normal exogenous



Platykurtic exogenous



Leptokurtic exogenous

Table 15. Linear Italian Model ($\chi^2_{(19)}$; $T = 200$)

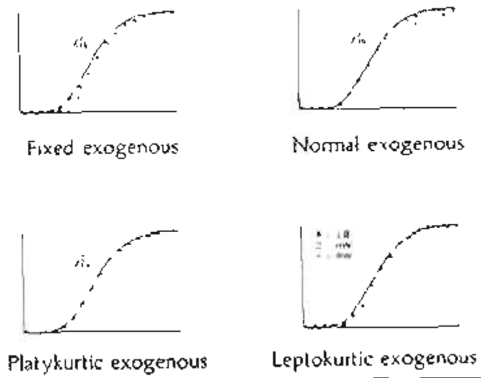


Table 16. Linear Italian Model ($\chi^2_{(19)}$; $T = 1000$)

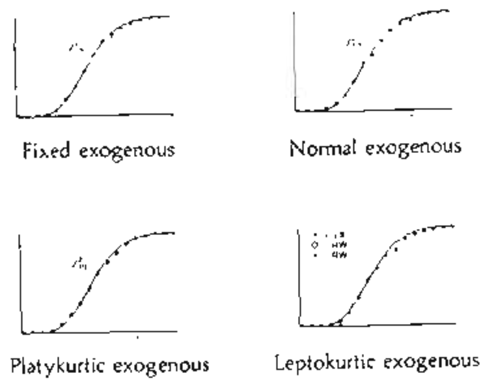


Table 17. Log-Linear Version of Klein's Model-I[a]

$$\begin{aligned}
\ln C_t &= a_1 + a_2 \ln P_t + a_3 \ln P_{t-1} \\
&\quad + a_4 \ln(W1 + W2)_t + u_{1,t} \\
I_t &= a_5 + a_6 P_t + a_7 P_{t-1} + a_8 K_{t-1} + u_{2,t} \\
W1_t &= a_9 + a_{10}(Y + T - W2)_t \\
&\quad + a_{11}(Y + T - W2)_{t-1} + a_{12} t + u_{3,t} \\
Y_t &= C_t + I_t + G_t - T_t \\
P_t &= Y_t - W1_t - W2_t \\
K_t &= K_{t-1} + I_t
\end{aligned}$$

Number of equations = 6
Number of stochastic equations $m = 3$
Number of structural unknown coefficients $n = 12$
Number of structural unknown parameters $n + m(m + 1)/2 = 18$

[a] The model, specifically designed for experimenting algorithms for nonlinear systems, is taken from Belsley (1980).

Table 18. Log-Linear Version of Klein's Model-I ($\chi^2_{(12)}$; $T = 42$)

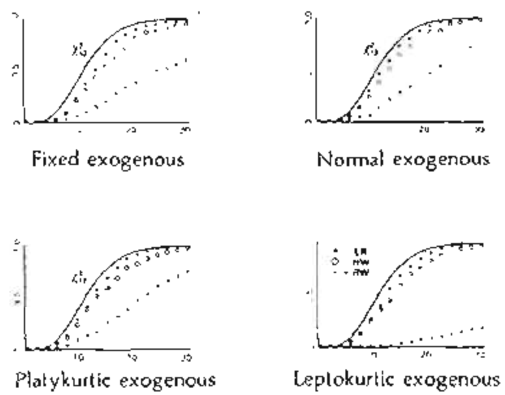


Table 19. Log-Linear Version of Klein's Model-I ($\chi^2_{(12)}$, T = 210)

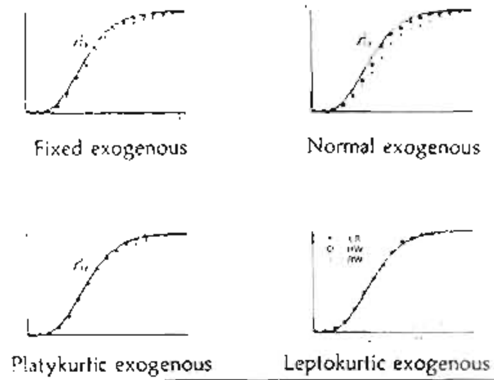
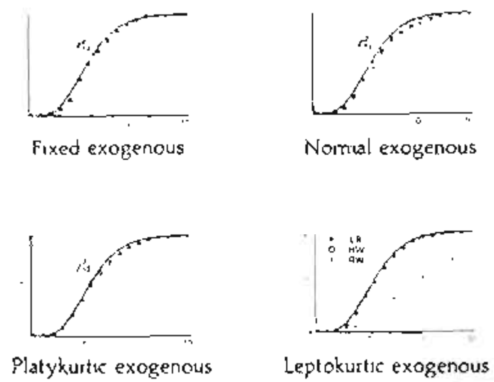


Table 20. Log-Linear Version of Klein's Model-I ($\chi^2_{(12)}$; T = 1050)



IX. CONCLUSIONS

In this chapter we have experimentally evaluated how costly it is to use the robust estimator of the coefficients covariance matrix in the Wald test on systems of simultaneous equations when there is no misspecification. We have investigated how fast the sampling distribution of the test converges to the asymptotic χ^2 distribution and have performed comparisons with the *likelihood ratio* test and with the traditional *Hessian* version of the *Wald* test.

The results have indicated that the cost may be very large depending on the values assumed by the explanatory variables of the model. The enormous deviations of the sampling distribution from the asymptotic χ^2 recommend a careful investigation of the problem of small sample correcting factors before the *robust Wald* test can be expected to have approximately the correct size for the sample lengths used in econometric applications.

ACKNOWLEDGMENTS

Some of the results described in this chapter were presented at the Fifth World Congress of the Econometric Society, M.I.T., Cambridge, MA, August 17–24, 1985. The authors are grateful to Gordon A. Hughes, Ingmar R. Prucha, and N. Eugene Savin for suggestions and comments. Responsibility, however, remains entirely with the authors.

NOTES

1. We first perform a least-squares estimation of the coefficients for all the equations of the model in such a way as to get a *reasonably good* starting point for the maximization process. Then we perform some iterations of FIML using a gradient algorithm based on a *generalized least-squares type* matrix that usually proves to be computationally more efficient in the first iterations (see Calzolari et al., 1987). Intermediate iterations are then made using a *Newton-like* algorithm based on the Hessian of the likelihood followed by a search of the maximum in the chosen direction. The last iterations are performed using Newton's method. An accurate computation of the maximum point is ensured by choosing a very tight convergence criterion: 10^{-9} as a relative tolerance on all coefficients in each Monte Carlo replication.

2. When dealing with maximum likelihood estimation, it is usual to *concentrate out* the σ parameters and then deal with the *concentrated* log-likelihood, which is only a function of the α parameters. The formulas for the first and second derivatives of the concentrated log-likelihood would be more complicated than those displayed in section III (for example, see Amemiya, 1977, p. 957 for the second derivatives), but would give several computational benefits. In particular, the dimensions of all the matrices involved in the computation would be $n \times n$ instead of $[n + m(m + 1)/2] \times [n + m(m + 1)/2]$. It can easily be proved that the Hessian of the concentrated log-likelihood is equal to $(A^{1,1})^{-1}$. This equality does not hold for the covariance estimators based on the outer products of the first derivatives of the log-likelihoods or of the concentrated log-likelihoods. However, Prucha (1984) proves that the equality holds again (algebraically, and not only in probability limit) for the first block of matrix C; in other words, C_{11} could also be obtained from Hessian and outer products of first-order derivatives of the concentrated log-likelihoods. There are, of course, applications of the *Wald* test where it is necessary to evaluate the covariance matrix of the entire set of estimated parameters (see, for example, Bhargava, 1987); the formulas of sections III and IV can be used also in these applications.

3. Of course, it would not be so if we were interested in getting very accurate measurements of the distributions in the critical region. In such a case accuracy would be helped not only on a larger number of replications, but also by the use of variance reduction algorithms. In particular, the control variate method proposed by Davidson and MacKinnon (1981) and applied by them (1983) to the *Lagrange multiplier* test on a linear regression model should be suitable. Also in the case of linear simultaneous

equations, we may calculate a control variate that has a known distribution in the small sample case, ℓ^*, and is at the same time, hopefully, strongly correlated with the Wald and LR test statistics. In sampling experiments such a variable may be obtained as an LM statistic, using the *true* values of the σ parameters in the score vector and using the *true* information matrix. This, however, seems not to be so simple if the system is nonlinear.

4. Whether or not the differences can be considered slight is certainly a matter of opinion. However, there is surely no doubt that they are very slight if compared with the differences that will be shown in some of the following experiments.

5. The extreme case of a *dummy* variable that is always zero except in one period cannot obviously be considered. This follows from Eq. (31), where k_t would be equal T, and from Chesher and Jewitt (1987, p. 1219); even more clearly, it follows from the *proposition* in Chesher and Jewitt (1984, p. 10).

REFERENCES

Amemiya, T. (1977) "The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous Equation Model." *Econometrica*, 45, 955–968.

Belsley, D. A. (1980) "On the efficient computation of the nonlinear full-information maximum-likelihood estimator." *Journal of Econometrics*, 14, 203–225.

Berndt, E. and N. E. Savin (1977) "Conflict among criteria for testing hypotheses in the multivariate linear regression model." *Econometrica*, 45, 1263–1278.

Bhargava, A. (1987) "Wald tests and systems of stochastic equations". *International Economic Review*, 28, 789–808.

Breusch, T. S. (1979) "Conflict among criteria for testing hypotheses: Extensions and comments." *Econometrica*, 47, 203–207.

Calzolari, G. and L. Panattoni (1984) "A simulation study on FIML covariance matrix." Pisa: Centro Scientifico IBM, discussion paper presented at the European Meeting of the Econometric Society, Madrid.

Calzolari, G. and L. Panattoni (1988) "Alternative estimators of FIML covariance matrix: A Monte Carlo study." *Econometrica*, 56, 701–714.

Calzolari, G., L. Panattoni, and C. Weihs (1987) "Computational efficiency of FIML estimation." *Journal of Econometrics*, 36, 299–310.

Chesher, A. and I. Jewitt (1984) "Finite sample properties of least square covariance matrix estimators." University of Bristol: Department of Economics, discussion paper No. 163.

Chesher, A. and I. Jewitt (1987) "The bias of a heteroskedasticity consistent covariance matrix estimator". *Econometrica*, 55, 1217–1222.

Davidson, R. and J. G. MacKinnon (1981) "Efficient estimation of tail-area probabilities in sampling experiments." *Economics Letters*, 8, 73–77.

Davidson, R. and J. G. MacKinnon (1983) "Small sample properties of alternative forms of the Lagrange multiplier test." *Economics Letters*, 12, 269–275.

Domowitz, I. and H. White (1982) "Misspecified models with dependent observations." *Journal of Econometrics*, 20, 35–58.

Engle, R. F. (1984) "Wald, likelihood ratio, and Lagrange multiplier tests in econometrics." In *Handbook of Econometrics*, Vol. 2, Z. Griliches and M. D. Intriligator (eds.), pp. 775–826. North-Holland, Amsterdam.

Evans, G. B. A. and N. E. Savin (1982) "Conflict among the criteria revisited: The W, LR and LM tests." *Econometrica*, 50, 737–748.

Gourieroux, C., A. Monfort, and A. Trognon (1984) "Pseudo maximum likelihood methods: Theory." *Econometrica*, 52, 681–700.

Klein, L. R. (1950) *Economic Fluctuations in the United States, 1921–1941*. John Wiley, New York. Cowles Commission Monograph No. 11.

MacKinnon, J. G. and H. White (1985) "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties." *Journal of Econometrics*, 29, 305–325.

Prucha, I. R. (1984) "On the estimation of the variance covariance matrix of maximum likelihood estimators in nonlinear simultaneous equation systems: A Monte Carlo study." University of Maryland, Department of Economics, Working paper No. 84-14.

Rettore, E. and U. Trivellato (1986) "Effetti dei dati provvisori sull'errore di previsione di modelli ad equazioni simultanee." In *Errori nei Dati Preliminari, Previsioni e Politiche Economiche*, a cura di U. Trivellato, pp. 267–284. CLEUP, Padova.

Rothenberg, T. J. (1973) *Efficient Estimation with A Priori Information*. Yale University Press, New Haven. Cowles Foundation Monograph No. 23.

Savin, N. E. (1976) "Conflict among testing procedures in a linear regression model with autoregressive disturbances." *Econometrica* 44, 1303–1315.

Sitzia, B. and M. Tivegna (1975) "Un modello aggregato dell'economia italiana 1952–1971." In *Contributi alla Ricerca Economica* No. 4, pp. 195–223. Banca d' Italia, Roma.

White, H. (1980) "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica*, 48, 817–838.

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

White, H. (1983) "Corrigendum." *Econometrica*, 51, 513.