L_1-Statistical Procedures and Related Topics
IMS Lecture Notes – Monograph Series (1997) Volume 31

A Lagrange multiplier approach to testing for serially dependent error terms

Hans Nyquist

University of Umeå, Sweden

Abstract: Testing for first-order auto-regressive errors in a linear regression model is considered. It is found that the L_1-norm based Lagrange multiplier test avoids computational difficulties caused by the dependency among the errors. Furthermore, the Lagrange multiplier test has the advantage that estimation of the error term density at zero is not required. As the error term variance increases and the error term density at zero becomes larger the asymptotic relative efficiency becomes more favorable for the L_1-norm based test relative to the corresponding least squares test.

Key words: Auto-regressive errors, L_1-norm estimation, Lagrange multiplier test.

AMS subject classification: 62J05, 62M10.

1 Introduction

The existence of serially dependent error terms when a linear regression model is used for analyzing time series data has attracted the attention of recent research. Elimination of estimation problems caused by the dependence involves both the detection of the dependence and the selection of an appropriate estimation technique for the model in question if dependence is found.

It is also well recognized in the literature that many data sets contain outliers or, alternatively, are well represented by distributions with fat tails. This has motivated the introduction of robust estimators, including the L_1-norm estimator. Simulation experiments of L_1-norm based estimators of models with serially dependent error terms are reported in e.g. Coursey and Nyquist (1983 and 1986). These experiments indicate that L_1-norm methods for estimation are preferable to least squares methods when the tails of the error distribution becomes fatter. The experiments also indicate that estimators that take account of the serially dependence outperform those estimators that ignore it.

Modelling for serially dependent errors results in a non-linear model, implying that standard techniques for computing estimates do not apply. One procedure for least squares estimation of linear regression models with serially dependent errors is that of Cochrane and Orcutt (1949). The Cochrane-Orcutt approach was extended to estimation using a more general criterion function in Coursey and Nyquist (1986) and further analyzed in Nyquist (1992). These analyses show that saddle points and multiple minima are very common when the L_1-norm criterion function is used. As a consequence, the L_1-norm based Cochrane-Orcutt procedure has a tendency to converge to points that do not define the global minimum of the criterion function. On the other hand, extensions of the Gauss-Newton technique to the L_1-norm case, as suggested by Osborne and Watson (1971) and Anderson and Osborne (1977), has shown more satisfactory results when it has been applied to models with serially dependent errors (Nyquist, 1992).

In this paper attention is restricted to the case with a first-order auto-regressive error process. The aim of the paper is to present and discuss the Lagrange multiplier test of the hypothesis of serially independent error terms. This test is particularly attractive from a computational viewpoint since it only requires estimation of the restricted model, which is in this case, a linear model. Computational difficulties caused by non-linearities are therefore avoided in this approach. Furthermore, the Lagrange multiplier test has the advantage that it does not require estimation of the error density at zero.

2 The model

We consider the linear regression model with first-order auto-regressive errors

$$y_t = x_t'\beta + u_t, \ t = 1, 2, \ldots, T, \tag{1}$$

$$u_t = \phi u_{t-1} + v_t, \ t = 1, 2, \ldots, T, \tag{2}$$

where y_t is a response variable, x_t is a p-vector of known regressors, β is a p-vector of unknown regression parameters, and u_t is the error generated by the auto-regressive process with parameter ϕ. The disturbances $v_1, v_2, \ldots, v_T$ are assumed to be independent and identically distributed.

Let X be the $T \times p$ matrix with x_t' as rows. For the subsequent analysis we assume that

Assumption 1: There exists a positive definite matrix D such that $\lim_{T\to\infty} T^{-1}X'X = D$.

Assumption 2: The common cumulative distribution function of the disturbances v_t, F, is differentiable at 0 with $F'(0) > 0$.

Assumption 3: The disturbances v_t have a finite variance, $d = V(v_t)$.

Our aim is to test the null hypothesis

$$H_0 : \phi = 0,$$

against the alternative

$$H_a : \phi \neq 0.$$

Lagging (1) by one period, multiplying by ϕ, and subtracting from (1) yields

$$y_t - \phi y_{t-1} = \left(x_t' - \phi x_{t-1}'\right)\beta + v_t, \;\; t = 2, 3, \ldots, T.$$

Conditioning on the first observation, the L_1-norm criterion function is defined as

$$S(\beta, \phi) = \sum_{t=2}^{T} \left| y_t - \phi y_{t-1} - \left(x_t' - \phi x_{t-1}'\right)\beta \right|$$

$$= \sum_{t=2}^{T} \left| y_t - h_t(\beta, \phi) \right|, \qquad (3)$$

where

$$h_t(\beta, \phi) = \phi y_{t-1} + \left(x_t' - \phi x_{t-1}'\right)\beta. \qquad (4)$$

The L_1-norm estimator $\left(\widehat{\beta}, \widehat{\phi}\right)$ is defined as the minimizer of $S(\beta, \phi)$. The computational problems caused by the serially dependency among the error terms stem from the non-linearity of $h_t(\beta, \phi)$.

The Gauss-Newton approach to minimize $S(\beta, \phi)$ utilizes a first-order Taylor approximation of $h_t(\beta, \phi)$ evaluated at a previous estimate $\left(\beta^{(s)}, \phi^{(s)}\right)$ of the parameters,

$$h_t^{(s)}(\beta,\phi) =$$

$$h_t\left(\beta^{(s)},\phi^{(s)}\right) + \left(\phi - \phi^{(s)}\right)\widehat{u}_{t-1}^{(s)} + \left(x_t' - \phi^{(s)}x_{t-1}'\right)\left(\beta - \beta^{(s)}\right), \quad (5)$$

where

$$\widehat{u}_t^{(s)} = y_t - x_t'\beta^{(s)}.$$

The minimizer of

$$S^{(s)}(\beta,\phi) = \sum_{t=2}^{T}\left|y_t - h_t^{(s)}(\beta,\phi)\right| \quad (6)$$

now yields new estimates, $\left(\beta^{(s+1)},\phi^{(s+1)}\right)$. The function $h_t^{(s)}(\beta,\phi)$ is linear in the parameters, so that standard routines, such as those of Barrodale and Roberts (1974) and Armstrong et al. (1979), can be used for finding the minimum of (6). Conditions for the convergence of the iteration process are found in Osborne and Watson (1971) and Anderson and Osborne (1977).

3 The Lagrange multiplier test

The Lagrange multiplier test is based on the gradient of the unrestricted estimation problem evaluated at the restricted estimate. Thus, under the null hypothesis that $\phi = 0$, the restricted model is the linear regression model ignoring dependent error terms. Denoting the restricted estimates by $\widehat{\beta}^{(r)}$ and using (5) we obtain the model

$$y_t = x_t'\beta + \phi\widehat{u}_{t-1}^{(r)} + u_t,\ t = 2,3,\ldots,T, \quad (7)$$

where

$$\widehat{u}_t^{(r)} = y_t - x_t'\widehat{\beta}^{(r)},\ t = 2,3,\ldots T$$

are the residuals computed from the restricted model. We find that the one step Gauss-Newton estimate of the auto-regressive parameter ϕ appears as the regression parameter for the constructed variable $\widehat{u}_{t-1}^{(r)}$, appearing in the linear model (7). The L_1-norm criterion function to be minimized is therefore

$$S_0(\beta,\phi) = \sum_{t=2}^{T}\left|y_t - x_t'\beta - \widehat{u}_{t-1}^{(r)}\phi\right|.$$

There are at least three immediate implications. First, the significance of including the auto-regressive error process to the model can be analyzed graphically by an added variable plot. Secondly, the significance of the auto-regressive process can formally be tested using the Lagrange multiplier test, by testing the significance of the regression parameter ϕ of the linear regression model (7). Thirdly, the model (7) can be estimated and the gradient for the Lagrange multiplier test can be computed by using standard routines designed for linear models.

The asymptotic distribution of L_1-norm estimators of parameters in linear regression models was derived by Bassett and Koenker (1978) and Amemiya (1982). An application of this theory to the L_1-norm estimator $\left(\widehat{\beta}, \widehat{\phi}\right)$ of the parameters of the linear model (7) yields that $\sqrt{T}\left(\widehat{\beta}' - \beta', \widehat{\phi}\right)'$ is asymptotically normally distributed with mean zero and covariance matrix

$$\omega^2 \begin{pmatrix} D^{-1} & 0 \\ 0 & d^{-1} \end{pmatrix},$$

where $\omega^2 = 1/\left(2f\left(0\right)\right)^2$, provided H_0 is true.

Following Koenker and Bassett (1982) we define

$$W\left(\delta_1, \delta_2\right) = \sum_{t=2}^{T} \left| u_t - \left(x_t'\delta_1 + \widehat{u}_{t-1}^{(r)}\delta_2\right)/\sqrt{T-1} \right|,$$

so that $W\left(\sqrt{T-1}\left(\widehat{\beta}^{(r)} - \beta\right), \sqrt{T-1}\left(\widehat{\phi} - \phi\right)\right) = S_0\left(\widehat{\beta}^{(r)}, \widehat{\phi}\right)$. We further define the normalized gradient of W as

$$g\left(\delta_1, \delta_2\right) = \frac{1}{\sqrt{T-1}} \sum_{t=2}^{T} \begin{pmatrix} x_t \\ \widehat{u}_{t-1}^{(r)} \end{pmatrix} sign\left(u_t - \left(x_t'\delta_1 + \widehat{u}_{t-1}^{(r)}\delta_2\right)/\sqrt{T-1}\right).$$

Evaluating g at $\left(\widehat{\delta}_1, \widehat{\delta}_2\right) = \left(\sqrt{T-1}\left(\widehat{\beta}^{(r)} - \beta\right), -\sqrt{T-1}\phi\right)$ yields the vector $g\left(\widehat{\delta}_1, \widehat{\delta}_2\right) = \left(\widehat{g}_1, \widehat{g}_2\right)'$, where

$$\widehat{g}_2 = \frac{1}{\sqrt{T-1}} \sum_{t=2}^{T} \widehat{u}_{t-1}^{(r)} sign\left(\widehat{u}_t^{(r)}\right).$$

If $\widehat{g}_2$ is large then H_0 is implausible. The Lagrange multiplier test statistic is now defined as the quadratic form of the gradient

$$\xi = \frac{\widehat{g}_2^2}{d} = \frac{\left\{\sum_{t=2}^{T} \widehat{u}_{t-1}^{(r)} sign\left(\widehat{u}_t^{(r)}\right)\right\}^2}{(T-1)\,d}$$

From Koenker and Bassett (1982) it follows that the distribution of ξ is approximately non-central χ^2 with one degree of freedom and non-centrality parameter $(T-1)\,\frac{d}{\omega^2}\phi^2$. In particular, when H_0 is true the asymptotic distribution of ξ is the central χ^2 distribution with one degree of freedom.

A comparison of the powers of the L$_1$-norm based Lagrange multiplier test and the corresponding test based on least squares can be done in terms of asymptotic relative efficiency. This quantity is the ratio of the noncentrality parameters of the limiting distributions, $ARE = d/\omega^2$. This comparison is therefore similar to a comparison of estimators efficiency. For normally distributed errors $ARE = 2/\pi \approx 0.64$ so that the least squares method is preferable. However, as the error term variance increases and ω^2 decreases, the L$_1$-norm based methods become more favorable in terms of ARE. At the Laplace distribution, for example, $ARE = 2$.

4 Final remarks

This paper describes a Lagrange multiplier test for testing for a first-order auto-regressive error process in a linear regression model. There are several immediate extensions of the procedure. First, since the variance d of the disturbances v_t is unknown in most of the applications, it needs to be replaced by an estimator in the definition of ξ. A possible estimator is obtained by first estimating ϕ by minimizing S_0 and than defining

$$\widehat{d} = \frac{1}{T-1}\sum_{t=2}^{T}\left(\widehat{u}_t^{(r)} - \widehat{\phi}\widehat{u}_{t-1}^{(r)}\right)^2 .$$

Secondly, Assumption 1 can be weakened to

Assumption 1': $\max_{1\le t\le T} x_t'\left(X'X\right)^{-1}x_t \to 0$, as $T\to\infty$.

This would change the normalization of $\widehat{\beta}-\beta$ from $\sqrt{T-1}$ to $\left(X'X\right)^{-1/2}$.

Thirdly, a test for a $p-th$ order auto-regressive error process

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \ldots + \phi_p u_{t-p} + v_t,\ \ t = 1,2,\ldots,T,$$

is obtained if (1) is lagged by 1, 2,..., p periods, multiplying the equations by ϕ_1, ϕ_2,..., ϕ_p, respectively, and subtracting from (1) to obtain

$$\begin{aligned} & y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \ldots - \phi_p y_{t-p} \\ = \ & \left(x_t' - \phi_1 x_{t-1}' - \phi_2 x_{t-2}' - \ldots - \phi_p x_{t-p}'\right)\beta + v_t,\ \ t = p+1, p+2,\ldots,T. \end{aligned}$$

The L$_1$-norm criterion function, given the first p observations, is now

$$S(\beta,\phi)=\sum_{t=p+1}^{T}|y_t-h_t(\beta,\phi)|$$

with

$$h_t(\beta,\phi)=$$
$$\phi_1 y_{t-1}+\phi_2 y_{t-2}+\ldots+\phi_p y_{t-p}+\left(x_t'-\phi_1 x_{t-1}'-\phi_2 x_{t-2}'-\ldots-\phi_p x_{t-p}'\right)\beta.$$

The restricted model under the hypothesis that $\phi_1=\phi_2=\ldots=\phi_p=0$ becomes

$$y_t=x_t'\beta+\phi_1\widehat{u}_{t-1}+\phi_2\widehat{u}_{t-2}+\ldots+\phi_p\widehat{u}_{t-p}+u_t,\ t=p+1,p+2,\ldots,T$$

and the Lagrange multiplier test of the hypothesis of no auto-regressive errors of order less than or equal to p is equivalent to the Lagrange multiplier test of the hypothesis that the regression coefficients ϕ_1, ϕ_2, ..., ϕ_p in the linear regression model are equal to zero.

Finally, it is worth pointing out that the approach for testing for autocorrelated error terms may be extended, under fairly mild regularity conditions, to the more general case of M-estimation of linear models.

References

[1] Amemiya, T. (1982). Two stage least absolute deviations estimators. *Econometrica* **50**, 689-712.

[2] Anderson, D.H. and Osborne, M.R. (1977). Discrete, nonlinear approximation problems in polyhedral norms. *Numer. Math* **28**, 143-156.

[3] Armstrong, R.D., Frome, E,L., and Kung, D.S. (1979). A revised simplex algorithm for the absolute deviation curve fitting problem. *Commun. Statist - Simul. Comput.* **8**, 175-190.

[4] Barrodale, I. and Roberts, F.D.K. (1974). Algorithm 478: Solution of an overdetermined system of equations in the ℓ_1-norm. *Commun. Assoc. Comput. Math.* **17**, 319-320.

[5] Bassett, G. and Koenker, R. (1978). Asymptotic theory of least absolute error regression. *J. Am. Statist. Assoc.* **73**, 618-622.

[6] Cochrane, D. and Orcutt, G.H. (1949). Applications of least squares regressions to relationships concerning autocorrelated error terms. *J. Am. Statist. Assoc.* **44**, 32-61.

[7] Coursey, D. and Nyquist, H. (1983) On least absolute error estimation of linear regression models with dependent stable residuals. *Rev. Econ. Statist.* **65**, 687-692.

[8] Coursey, D. and Nyquist, H. (1986). A procedure for obtaining M-estimates in regression models with serially dependent errors. *J. Time Ser. Analys.* **7**, 255-267.

[9] Koenker, R. and Bassett, G. (1982). Tests of linear hypotheses and ℓ_1 estimation. *Econometrica* **50**, 1577-1583.

[10] Nyquist, H. (1992). L_1-norm estimation of regression models with serially dependent error terms. In *L_1 Statistical Analysis and Related Methods*, Ed. Y. Dodge, pp. 253-264. Amsterdam: North-Holland.

[11] Osborne, M.R. and Watson, G.A. (1971). On an algorithm for discrete nonlinear L_1 approximation. *Comput. J.* **14**, 184-188.