Improved estimation of the linear regression model with autocorrelated errors

A Chaturvedi
University of Allahabad

Tran Van Hoa
University of Wollongong

Ram Lal
University of Allahabad

Recommended Citation
http://ro.uow.edu.au/commwpapers/327
IMPROVED ESTIMATION OF THE LINEAR REGRESSION MODEL WITH AUTOCORRELATED ERRORS

A. Chaturvedi  
*Department of Mathematics and Statistics*  
*University of Allahabad*  
*Allahabad, India*

Tran Van Hoa  
*Department of Economics*  
*The University of Wollongong*  
*Wollongong NSW 2500 Australia*

and

Ram Lal  
*Department of Statistics*  
*University of Allahabad*  
*Allahabad, India*

Working Paper 90-14
ABSTRACT

For estimating the coefficient vector of a linear regression model with first order autoregressive disturbances, a family of Stein-rule estimators based on the two-step Prais-Winsten estimating procedure is considered and an Edgeworth type asymptotic expansion for its distribution is derived. The performance of this family of estimators relative to the two-step Prais-Winsten estimator is also derived under a squared error loss function.

Keywords: Linear regression model, autocorrelated errors, squared error loss function, improved estimators, Stein estimators, Prais-Winsten estimators, Edgeworth expansions, generalised least squares estimators.

Acknowledgement. The first author gratefully acknowledges the research grant from the Department of Economics, University of Wollongong, Wollongong, Australia, during his visit there in 1989.
1. INTRODUCTION

For estimating the coefficient vector of a linear regression model with disturbances following a first order autoregressive scheme, several estimators have been analysed with the help of empirical methods [see for example Kramer (1980), Maeshiro (1976), Park and Mitchell (1980) and Spitzer (1979)]. Rothenberg (1984) considered the case when the error covariance matrix depends on a finite number of parameters, and derived the approximate expression for the distribution of the two-step generalised least squares (GLS) estimator. Magee et. al. (1985) considered several classes of two-step Cochrane-Orcutt and Prais-Winsten estimators which arise from the choice of the various estimated autocorrelation coefficients and obtained the large sample expressions of the dispersion matrices, and analysed the efficiency properties of the estimators with the help of a numerical experiment. Magee (1985) also adopted a general method to derive Nagar expansions for iterative estimators and applied it to obtain the approximate dispersion matrices for the iterated Prais-Winsten and maximum likelihood estimators.

While several families of improved or shrinkage estimators with superior properties in terms of a quadratic loss function have been developed for the linear regression model with i.i.d. disturbances [see Judge and Bock (1978) and Vinod and Ullah (1981)], no study dealing with improved estimation in the case of autocorrelated disturbances has been reported. In this paper, we present an attempt in this direction and consider a general family of improved or shrinkage estimators based on the two-step Prais-Winsten estimator for the coefficient vector of a linear regression model with disturbances generated by a first order autoregressive scheme. Edgeworth type asymptotic expansions for the distribution of the proposed family of estimators are derived and the risk under a quadratic loss function of the
resulting estimators is compared with that of the two-step Prais-Winsten estimator.\(^1\) The results of a numerical experiment are also reported.

2. THE MODEL AND THE FAMILY OF ESTIMATORS

Consider the linear regression model:

\[ y = X\beta + u \]  \hspace{1cm} (1)

where \( y \) is a \( T \times 1 \) vector of observations on the dependent variable, \( X \) is a \( T \times p \) matrix of observations on \( p \) independent variables, \( \beta \) is a \( p \times 1 \) vector of unknown regression coefficients and \( u \) is a \( T \times 1 \) disturbance vector. The elements of \( u \) are assumed to follow an AR(1) process:

\[ u_t = \rho u_{t-1} + \epsilon_t \hspace{0.5cm} (t = 1, 2, \ldots, T) \]

where \( \rho \) (\( |\rho| < 1 \)) is the unknown autocorrelation coefficient. Further

\[
E(\epsilon_t) = 0 \\
E(\epsilon_t \epsilon_t' + s) = \begin{cases} \psi & \text{if } s = 0 \\
0 & \text{otherwise.} \end{cases}
\]

Thus \( E(u) = 0 \) and \( E(u u') = \sigma^2 \Sigma \) where \( \sigma^2 \frac{\psi}{(1 - \rho^2)} \) and \( \Sigma \) is a \( T \times T \) matrix with \((i,j)\)-th element as \( \rho^{i-j} \).

The ordinary least squares (OLS) estimator of \( \beta \) is given by

\(^1\) The results presented in this paper apply equally well to Stein rule estimator based on any feasible GLS estimator involving a first order efficient estimator of \( \rho \), e.g., two step or iterated Prais-Winsten estimators, the maximum likelihood estimator or the estimators derived from the Durbin-Watson statistic.
\[ b = (X'X)^{-1} X'y \]

while the GLS estimator is

\[ \hat{\beta}^* = (X'P'PX)^{-1} X'P'Py, \]

where \( P \) is \( T \times T \) triangular matrix with (i,j)-th element \((1 - \rho^2)^{\frac{1}{2}}\) if \( i = j = 1 \), \( 1 \) if \( i = j = 2,3,\ldots,T \), \(-\rho\) if \( i = j + 1 = 2,3,\ldots,T \) and 0 otherwise.

Since \( \rho \) is usually unknown, Prais and Winsten (1954) suggested the replacement of \( \rho \) by its consistent estimator\(^2\)

\[
\hat{\rho} = \frac{\sum_{t=1}^{t-1} \hat{u}_t \hat{u}_{t+1}}{\sum_{t=1}^{T} \hat{u}_t^2}
\]

where \( \hat{u}_t \) is the t-th element of the vector \( \hat{u} = y - Xb \). They therefore obtained the following estimator of \( \beta \):

\[
\hat{\beta} = (X' \hat{\rho} \cdot \hat{\rho} X)^{-1} X' \hat{\rho} \cdot \hat{\rho} y
\]

where \( \hat{\rho} \) is obtained by replacing \( \rho \) by \( \hat{\rho} \) in \( \rho \).

Now we define the following family of improved or shrinkage estimators for \( \beta \):

---

\(^2\) The original estimator of \( \rho \) proposed by Prais and Winsten involved in the denominator sum of squares of \( \hat{u}_t \)'s from \( t \) equal to 2 to \( T-1 \). However, the adjustment in the denominator made here does not affect the results in the paper which only depend on \( \hat{\rho} \) to order \( O_p(T^{-\frac{1}{2}}) \) whereas the adjustment affects \( \hat{\rho} \) to order \( O_p(T^{-1}) \).
\[ \hat{\beta}(k) = 1 - k \frac{1}{T} \left[ \frac{(y - X\hat{\beta})' \hat{\beta} - \hat{\beta} (y - X\hat{\beta})}{\hat{\beta} \cdot x' \hat{\beta} - \hat{\beta} \cdot x \hat{\beta}} \right] \hat{\beta} \] (4)

Obviously for \( k = 0 \), \( \hat{\beta}(k) \) reduces to \( \hat{\beta} \).

In the following section, we study the asymptotic distribution of \( \hat{\beta}(k) \).

3. ASYMPTOTIC DISTRIBUTION OF THE FAMILY OF ESTIMATORS

Let us introduce the following notation:

\[ \Omega = T(X'\Sigma^{-1}X)^{-1}, \quad Q = \Sigma - \frac{1}{T} X \Omega X', \]
\[ \theta = \beta', \quad M = 1_T - X(X'X)^{-1}X', \]
\[ \mu = -\frac{k \sigma}{\theta T^2} \Omega^{\frac{1}{2}} \beta, \]
\[ V = I_T - \frac{1 - \sigma^2}{\theta^2 T^2} \Omega^{\frac{1}{2}} X' Q X \Omega^{\frac{1}{2}} + \frac{2k \sigma^2}{\theta^2 T} (2 \Omega^{\frac{1}{2}} \beta' \Omega^{\frac{1}{2}} - \theta 1_T), \]
\[ \pi(k) = \frac{T^\frac{1}{2}}{\sigma} \Omega^{\frac{1}{2}} [\hat{\beta}(k) - \beta]. \]

**THEOREM:** Let the matrix \( \left( \frac{1}{T} X'X \right) \) tends to a finite nonsingular matrix as \( T \) tends to infinity, \( |P| < 1 \) and the coefficient vector \( \beta \) is non-null. Then the asymptotic distribution of \( \pi(k) \), to order \( O(T^{-1}) \), is normal with mean vector \( \mu \) and dispersion matrix \( V \).
Using the above theorem, we can easily obtain the following approximate expressions for the bias vector, to order $0(T^{-1})$, and the mean squared error matrix, to order $0(T^{-2})$, of $\hat{\beta}(k)$ as:

\[ E[\hat{\beta}(k) - \beta] = -\frac{k\sigma^2}{\theta T} \]

(5)

\[ E[\hat{\beta}(k) - \beta] [\hat{\beta}(k) - \beta]' = \frac{\sigma^2}{T} \left[ \Omega + \frac{(1-p^2)}{p^2 T^2} \Omega X'QX + \frac{k\sigma^2}{\theta^2 T} \{(4 + k) \beta\beta' - 2 \theta \Omega \} \right] \]

(6)

Obviously, if we take $k = 0$ in (6), we get the asymptotic expression for the mean squared error matrix of $\hat{\beta}$ [see Ullah et. al. (1983)].

Consider the quadratic loss function $L(\tilde{\beta}) = (\tilde{\beta} - \beta)' C (\tilde{\beta} - \beta)$ where $\tilde{\beta}$ is an arbitrary estimator of $\beta$, $C$ is a $p \times p$ symmetric, positive definite matrix, and denote the risk associated with $\tilde{\beta}$ by $R[\tilde{\beta}]$. Then, to order $0(T^{-2})$, the approximate expressions for $R[\hat{\beta}(k)]$ is given by:

\[ R[\hat{\beta}(k)] = \frac{\sigma^2}{\theta T} \left[ \text{tr} (\Omega C) + \frac{(1-p^2)}{p^2 T^2} \text{tr} (\Omega X'QX \Omega C) \right. \]

\[ + \frac{k\sigma^2}{\theta^2 T} \left. \{(4 + k) \frac{\beta'CV\beta}{\theta} - 2 \text{tr} (\Omega C)\} \right] \]

(7)

Let $\lambda(.)$ be the maximum characteristic root of the matrix inside the bracket and

\[ d = \frac{\text{tr} (\Omega C)}{\lambda (.)}. \]
Since \( (\beta' C \beta / \theta) \leq \lambda (\Omega C) \), a sufficient condition for the difference between the risks of \( \hat{\beta} \) and \( \hat{\beta}(k) \), to order \( O(T^{-2}) \), to be non-negative is given by

\[
0 \leq k \leq 2 (d - 2).
\]

In particular, if we take \( C = \Omega^{-1} \), the expression (7) reduces to

\[
R[\hat{\beta}(k)] = \frac{\sigma^2}{T} \left[ p + \frac{1-p^2}{\rho^2 T^2} \text{tr} (X'QX\Omega) + \frac{k \sigma^2}{\theta T} \{(4 + k) - 2p\} \right]
\]

and the dominance condition (8) becomes

\[
0 \leq k \leq 2 (p - 2) ; p > 2.
\]

It can be easily verified that the optimum value of \( k \) obtained by minimizing expression (9) for \( R[\hat{\beta}(k)] \) is \( p - 2 = k_0 \) (say). For this optimum value \( k_0 \) of \( k \), the relative gain in efficiency of \( \hat{\beta}(k_0) \) over \( \hat{\beta} \), to order \( O(T^{-1}) \), is given by

\[
\Delta = \frac{R[\hat{\beta}] - R[\hat{\beta}(k_0)]}{R[\hat{\beta}]}
\]

\[
= \frac{(p - 2)^2 \sigma^2 (1 - p^2)}{T \delta (1 + \rho^2 - 2 \gamma p)}
\]

where \( \delta = (\beta'X'X\beta / T) \), \( \gamma = (\beta'X'DX\beta / 2(\beta'X'X\beta)) \) and \( D \) is a \( T \times T \) matrix with \( (i,j) \)-th element 1 if \( i = j \pm 1 \) and 0 otherwise; \( i,j=1,2,...T \).

4. A SIMULATION STUDY
In the theorem above, the expansion for the distribution of $R[\hat{\beta} (k)]$ is derived for the case of $\beta$ being a non-null vector. Thus (4) may not be a reasonable approximation for $R[\hat{\beta} (k)]$ for any given $T$, no matter how large, and for $\beta$ in the neighbourhood of $0$. As a result, the derivation of the relative gain in efficiency of $\beta(k_0)$ over $\beta$ in this special case is analytically difficult if not intractable. To circumvent the problem, we have instead used a Monte Carlo simulation to investigate the relative gain of our estimator.

In the simulation, we formulate a total of 110 linear regression models with autocorrelated disturbances. Each model differs from the other in terms of the true values of $\beta$ and the magnitudes of the autocorrelation coefficient $\rho$. For $\beta$, their values are significant only at the fourth or higher digit after the decimal point. For $\rho$, the values range from -1.00 to 1.00 with an increment of 0.20.

For $\beta$, the values are based on those used by Chi and Judge (1985) in their simulation study on improved estimators. The data of the independent variables are obtained from an early Monte Carlo study by Kmenta and Gilbert (1968). The disturbance variance $\sigma^2$ is arbitrarily set equal to 100 for all cases. The simulation results on the relative gain in efficiency of $\beta(k_0)$ over $\beta$ in all 110 linear regression models are given in Table 1. The relative gain is denoted by $R(ML/S)$ and defined as $100[MSE(\beta)/MSE(\beta_0) -1]$, where $MSE(\beta)$ is the mean squared errors for the estimator $\beta$, and similarly for $\beta_0$. In all linear regression models, the relative gain is computed for illustration from 100 statistical trials.

The general conclusion that emerges from the simulation results reported in Table 1 appears to indicate the uniform gain in efficiency of the
estimator $\hat{\beta}(k_0)$ over the estimator $\beta$ in all 110 linear regression models in which $\beta$ is in the neighbourhood of zero. Thus, the expansion for the distribution of $R[\hat{\beta}(k)]$ as given in the theorem appears to be valid generally. As expected however, the relative gain, while dependent on the value of $p$, does not appear to be a monotonously increasing or decreasing function of $p$. The exact relationship between the gain and $k$, $\sigma^2$, and $T$ is given in (11).

5. DERIVATION OF THE RESULTS

Suppose $J$ is a diagonal matrix with first and last diagonal elements equal to one and remaining diagonal elements equal to $(1 - p^2)$, and $B$ is a $T \times T$ symmetric matrix with $(i,j)$th element $-p$ if $i = j$, $\frac{1}{p(1-p^2)}$ if $i = j \pm 1$ and 0 otherwise.

Following Ullah et. al. (1983), to order $O(T^{-1})$, $\pi(k)$ can be written as:

$$
\pi(k) = \frac{\sqrt{T}}{\sigma} \Omega^{\frac{1}{2}} \left[ \hat{\beta} - \beta - \frac{(1-p^2)\sigma^2 + \frac{1}{T} (y-X\hat{\beta})' \hat{\beta} (y-X\hat{\beta}) - (1-p^2)\sigma^2}{\hat{\beta}'X'I\hat{\beta}} \right],
$$

$$
= \frac{\sqrt{T}}{\sigma} \Omega^{\frac{1}{2}} \left[ \zeta_{-1} + \zeta_1 + \zeta_{-\frac{1}{2}} - \frac{k\sigma^2}{\theta_T} \left\{ \beta + \frac{1}{T} \Omega X'\Sigma^{-1} u - \frac{2}{T} (\beta'X'\Sigma^{-1} u)\beta \right\} \right],
$$

where

$$
\zeta_{-\frac{1}{2}} = \frac{1}{T} \Omega X'\Sigma^{-1} u
$$

$$
\zeta_{-1} = -\frac{\theta_{-\frac{1}{2}}}{P(1-P^2)T} \Omega X'JQ\Sigma^{-1} u
$$
\[ \zeta^0 = \frac{1}{\rho(1-\rho^2)T} \Omega X' \left[ \theta_1^2 \left\{ I_T - \frac{1}{(1-\rho^2)T} JX'X'J \right\} - \rho \theta_1 J \right] \Omega^{-1} u \]

with

\[ \theta_1 = \frac{1}{T\sigma^2} u'MBMu \]

\[ \theta_1 = -\frac{1}{T\sigma^4} u'MBMu \left( \frac{1}{T} u'Mu - \sigma^2 \right) \]

Now, for any \( p \times 1 \) vector \( h \) the cumulant generating function of \( \pi(k) \), to order \( O(T^{-1}) \), is given by

\[ K(h) = E \left[ \exp \{ ih'\pi(k)\} \right] \]

\[ = -i \frac{k\sigma}{\theta \sqrt{T}} (h'\Omega^{-1} \beta) + \log E \left[ \exp \left( \frac{i}{\sigma \sqrt{T}} h'\Omega^{1/2} X'\Sigma^{-1} u \right) \right] \]

\[ \left\{ 1 + \frac{i}{\sigma} (h'\Omega^{-1} \zeta) - \frac{T}{2\sigma^2} (h'\Omega^{-1} \zeta)^2 \right\} + \frac{i}{\sigma} (h'\Omega^{-1} \beta) - \frac{i}{\theta} h'h + \log \left[ 1 - \frac{(1-\rho^2)}{2\rho^2 T^2} \right] \]

\[ (h'\Omega^{1/2} X'QX\Omega^{1/2} h) + \frac{k\sigma^2}{\theta T} (h'h - \frac{2}{\theta} h'\Omega^{-1/2} \beta \beta' \Omega^{-1/2} h) \]

\[ = i h' \mu - \frac{i}{2} h'Vh + O(T^{-3}). \]

Now using the inversion formula
\[ f(\pi) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp \left[ K(h) - ih\pi \right] dh, \]

where \( f(\pi) \) denotes the pdf of \( \pi(k) \), we obtain the results of the theorem.
REFERENCES


