Estimation of linear regression models with AR(1) errors *

Jean-Marie Dufour †
McGill University

First version: March 2002
Revised: March 2008
This version: March 31, 2008
Compiled: March 31, 2008, 2:35pm

* This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec), and a Killam Fellowship (Canada Council for the Arts).

† William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

Contents

1. Model

Consider the model

$$y_t = x_t'\beta + u_t\ , \quad t = 1, \ldots, T \tag{1.1}$$

where

$$u_t = \rho u_{t-1} + \varepsilon_t\ , \quad t = \ \ldots\ ,\ 0,\ 1,\ 2,\ \ldots \tag{1.2}$$

$$|\rho| < 1\ , \tag{1.3}$$

$$\{\varepsilon_t\}_{t=1}^{T} \text{ is a sequence of i.i.d. disturbances,} \tag{1.4}$$

$$\mathsf{E}\,(\varepsilon_t) = 0\ , \quad \mathsf{V}\,(\varepsilon_t) = \sigma^2\ , \forall t\ . \tag{1.5}$$

Under these hypotheses,

$$\mathsf{V}\,(u_t) = \frac{\sigma^2}{1-\rho^2}\ , \forall t\,,$$

and, in particular,

$$\mathsf{V}\,(u_1) = \frac{\sigma^2}{1-\rho^2}\ .$$

2. Transformed model

Model (1.1) can be transformed as follows:

$$y_t^* = x_t^*\beta + u_t^*\ , \quad t = 1, \ldots, T \tag{2.1}$$

where

$$\begin{aligned} y_t^* &= y_t\,(\rho) = \begin{cases} \sqrt{1-\rho^2}\; y_1\ , & t = 1 \\ y_t - \rho y_{t-1}\ , & t = 2, \ldots, T \end{cases} \\ x_t^* &\equiv x_t\,(\rho) = \begin{cases} \sqrt{1-\rho^2}\; x_1\ , & t = 1 \\ x_t - \rho x_{t-1}\ , & t = 2, \ldots, T \end{cases} \\ u_t^* &= \begin{cases} \sqrt{1-\rho^2}\; u_1\ , & t = 1 \\ \varepsilon_t\ , & t = 2, \ldots, T \end{cases} \end{aligned}$$

i.e.

$$\begin{cases} \sqrt{1-\rho^2}\; y_1 = \sqrt{1-\rho^2}\; x_1'\beta + \sqrt{1-\rho^2}\; u_1\ , & \\ y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})'\,\beta + \varepsilon_t\ , & t = 2, \ldots T\ . \end{cases} \ . \tag{2.2}$$

3. Estimation

The method of generalized least squares (with known ρ) leads one to minimize with respect to β the function:

$$\begin{aligned} S_1(\rho, \beta) &= \Sigma_{t=1}^{T} \left[y_t(\rho) - x_t(\rho)' \beta \right]^2 \\ &= \left(1 - \rho^2\right) \left(y_1 - x_1' \beta\right)^2 + \Sigma_{t=2}^{T} \left[(y_t - \rho y_{t-1}) - (x_t - \rho x_{t-1})' \beta \right]^2 \ . \end{aligned}$$

The main problem here comes from the fact that ρ is generally unknown and must be estimated. We shall distinguish between two groups of methods to estimate this model:

1. approximate generalized least squares
   (a) without correction for the first observation,
   (b) with correction for the first observation;
2. maximum likelihood (ML).

3.1. Approximate generalized least squares

This method involves two main steps:

1. we estimate ρ by an appropriate estimator $\hat{\rho}$;
2. we estimate β by minimizing $S_1(\hat{\rho}, \beta)$ with respect to β [or by minimizing an approximation of $S_1(\hat{\rho}, \beta)$].

From there, we can distinguish between types of approaches depending on whether the first observation is taken to be fixed or random.

3.1.1. Methods where the first observation is taken as given

To avoid the complications entailed by applying a special transformation to the first observation, we consider the regression:

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})' \beta + \varepsilon_t \ , \quad t = 2, \ldots, T \ . \tag{3.1}$$

In other words, y_1 is taken as given (we condition on y_1). For ρ known (or given), we can estimate β by applying ordinary least squares (OLS) to (3.1), which is equivalent to solving the problem:

$$\min_{\beta} S(\rho, \ \beta)$$

where

$$S(\rho,\beta) = \Sigma_{t=2}^{T}\left[y_t(\rho) - x_t(\rho)'\beta\right]^2$$
$$= \Sigma_{t=2}^{T}\left[(y_t - \rho y_{t-1}) - (x_t - \rho x_{t-1})'\beta\right]^2 \ .$$

Similarly, for β known, we can estimate ρ by estimate the equation:

$$y_t - x_t'\beta = \left(y_{t-1} - x_{t-1}'\beta\right)\rho + \varepsilon_t \ , \quad t = 2, \ldots, T \ , \tag{3.2}$$

which is equivalent to solving the problem:

$$\min_{\rho} S(\rho,\beta) \ .$$

The only remaining problem now consists in estimating ρ. This will be done by solving the problem:

$$\min_{\rho,\beta} S(\rho,\beta) \ .$$

Several algorithms have been proposed to do this.

3.1.1.1. Hildreth-Lu algorithm .

1. Determine a grid for $-1 < \rho < 1$, *e.g.*,

   $$\rho = -0,95, -0,90, \ \ldots \ , \ 0,90, 0,95$$

   or

   $$\rho = -0,99, -0,98, \ \ldots \ , \ 0,98, 0,99 \, .$$

2. For each ρ, estimate (3.1) par OLS:

   $$\min_{\beta} S(\rho,\beta) \to \hat{\beta} = \hat{\beta}(\rho) \ .$$

3. Choose the value of ρ which minimizes $S(\rho,\beta)$.

3.1.1.2. Cochrane-Orcutt algorithm .

1. Choose an initial value for ρ : $\hat{\rho}_0$. For example, we can estimate β by OLS on the equation

   $$y_t = x_t'\beta + u_t$$

which yields $\hat{\beta}_0$, and then estimate ρ by

$$\hat{\rho}_0 = \Sigma_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} / \Sigma_{t=1}^{T} \hat{u}_t^2$$

where

$$\hat{u}_t = y_t - x_t' \hat{\beta}_0 \ , \quad t = 1, \ldots, T \ .$$

2. Estimate (3.1) with $\rho = \hat{\rho}_0 \to \hat{\beta}_1$.
   Estimate (3.2) with $\beta = \hat{\beta}_1 \to \hat{\rho}_1$.
   Estimate (3.1) with $\rho = \hat{\rho}_1 \to \hat{\beta}_2$.
   Estimate (3.2) with $\beta = \hat{\beta}_2 \to \hat{\rho}_2$.
   Etc.

3. We stop when ρ changes by less than a certain percentage δ (the tolerance of the algorithm), *e.g.*, $\delta = 0.001$.

3.1.1.3. Nonlinear estimation (Gauss-Newton) . The equation (1.1), for $t = 2, \ldots, T$, can be written:

$$\begin{aligned} y_t &= x_t' \beta + u_t \\ &= x_t' \beta + \rho u_{t-1} + \varepsilon_t \\ &= x_t' \beta + \rho \left(y_{t-1} - x_{t-1}' \beta \right) + \varepsilon_t \ , \quad t = 2, \ldots, T \ . \end{aligned} \tag{3.3}$$

The estimation of this equation can be viewed as a problem in nonlinear regression. Any appropriate algorithm (for example, the Gauss-Newton algorithm) can be used to estimate the model and thus ρ.

3.1.1.4. Durbin's method . This method is based on the following reparametrization of (3.3) :

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})' \beta + \varepsilon_t \ , \quad t = 2, \ldots, T$$

hence

$$\begin{aligned} y_t &= \rho y_{t-1} + (x_t - \rho x_{t-1})' \beta + \varepsilon_t \\ &= \rho y_{t-1} + x_t' \beta - x_{t-1}' (\rho \beta) + \varepsilon_t \\ &= \rho y_{t-1} + x_t' \beta + x_{t-1}' \beta_1 + \varepsilon_t \ , \quad t = 2, \ldots, T \ . \end{aligned} \tag{3.4}$$

If we estimate (3.4) by OLS, the estimated coefficient for y_{t-1}, say r_1, is un estimator of ρ.

3.1.2. Methods where the first observation is taken into account

After estimating ρ, we estimate β by setting $\rho = \hat{\rho}$ in (2.1) [or (2.2)], *i.e.* we minimize (with respect to β) the function:

$$\begin{aligned} S_1(\hat{\rho}, \beta) &= \left(1-\hat{\rho}^2\right)(y_1 - x_1'\beta)^2 + \Sigma_{t=2}^{T}\left[(y_t - \hat{\rho}y_{t-1}) - (x_t - \hat{\rho}x_{t-1})'\beta\right]^2 \\ &= \left(1-\hat{\rho}^2\right)(y_1 - x_1'\beta)^2 + S(\hat{\rho}, \beta)\ . \end{aligned}$$

This method is often called the "Prais-Winsten method". There are as many variants of this method as there are methods for estimating ρ. In particular, we can estimate ρ :

1. by applying any of the methods described in 3.1.1;
2. by minimizing $S_1(\rho, \beta)$ with respect to ρ and β simultaneously [using a grid or a nonlinear optimization algorithm].

3.2. Maximum likelihood

If we suppose that the ε_t are i.i.d. $N[0, \sigma^2]$, the likelihood function of $y_1, \ldots, y_T$ is

$$L \equiv L\left(y_1,\ \ldots\ ,\ y_T; \beta, \rho, \sigma^2\right) = \frac{\sqrt{1-\rho^2}}{\left(2\pi\sigma^2\right)^{T/2}} \exp\left\{-\frac{S_1(\rho, \beta)}{2\sigma^2}\right\} \tag{3.5}$$

and the log-likelihood is

$$\ln L = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 + \frac{1}{2}\ln\left(1-\rho^2\right) - \frac{S_1(\rho, \beta)}{2\sigma^2}\ . \tag{3.6}$$

The maximum likelihood estimator can be obtained by maximizing $\ln L$ with respect to ρ, β and σ^2. If $\hat{\rho}$ and $\hat{\beta}$ are ML ML estimators of ρ and β, we see easily that

$$\hat{\sigma}^2 = \frac{1}{T} S_1\left(\hat{\rho}, \hat{\beta}\right)\ .$$

If we replace σ^2 by $\frac{1}{T}S_1(\rho, \beta)$, we see that we can obtain ρ and β by minimizing (with respect to ρ and β) the function:

$$\begin{aligned} \ln(L^*) &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\left[\frac{1}{T}S_1(\rho, \beta)\right] + \frac{1}{2}\ln\left(1-\rho^2\right) - \frac{T}{2} \qquad (3.7) \\ &= -\frac{T}{2}\left[\ln(2\pi) + 1 + \ln\left(\frac{1}{T}\right)\right] - \frac{T}{2}\ln\left[S_1(\rho, \beta)\right] + \frac{1}{2}\ln\left(1-\rho^2\right)\ . \end{aligned}$$