Random Coefficient Autoregressive Models for Panel Data

Markku Rahiala
University of Oulu, Department of Mathematical Sciences
Linnanmaa
FIN-90570 Oulu, Finland
Markku.Rahiala@oulu.fi

1. Definition of the model

The use of autoregressive models with exogenous covariates for panel data is well established. See for instance Holz-Eakin, Newey and Rosen (1988), Arellano (1989), Arellano and Bond (1991) and also Rosner et al. (1985). If the shapes of individual trajectories vary considerably, the ARX-model can be generalised to allow for individual parameters for each cross-sectional unit. If these individual parameters are taken as Gaussian random variables, the model can be written as

$$y_{it} = \mu_i + \phi_{i1} y_{i,t-1} + ... + \phi_{ip} y_{i,t-p} + \beta_i' X_{it} + \varepsilon_{it} \ , \qquad \varepsilon_{it} \sim N(0, \sigma^2) \tag{1}$$

where $X_{it} = (\ x_{i1t} \quad ... \quad x_{imt}\ )'$ are exogenous, $\theta_i = (\ \mu_i \quad \phi_{i1} \quad .... \quad \phi_{ip} \quad \beta_i'\ )' \sim N_{m+p+1}(\theta, \Sigma)$, $\theta = (\ \mu \quad \phi_1 \quad .... \quad \phi_p \quad \beta'\ )'$ and $\{\varepsilon_{it}\}$ and $\{\theta_i\}$ are mutually independent and independent of $\{X_{it}\}$. Quite often only some of the components of θ_i will be assumed stochastic. To avoid singularity of Σ we write θ_i in the form

$$\theta_i = Q\psi_i + \theta \quad , \quad \psi_i \sim N_r(0, \sigma^2 \Sigma_o) \qquad (r \leq p + m + 1),$$

where Σ_o is non-singular. By defining $X_{it}^* = (\ 1 \quad y_{i,t-1} \quad .... \quad y_{i,t-p} \quad X_{it}'\ )'$, $X_i^* = (\ X_{i,p+1}^* \quad .... \quad X_{i,n_i}^*\ )'$, $S_i = (\ y_{i,1} \quad .... \quad y_{i,p}\ )'$, $Y_i = (\ y_{i,p+1} \quad .... \quad y_{i,n_i}\ )'$, $X_i = (\ X_{i,p+1} \quad .... \quad X_{i,n_i}\ )'$ and $Z_i = X_i^* Q$ we can write the whole model (1) in the more compact form

$$Y_i = X_i^* \theta + Z_i \psi_i + \varepsilon_i \quad , \quad \varepsilon_i \sim N_{n_i}(0, \sigma^2 I) \quad , \quad i = 1, ..., N \ . \tag{2}$$

The distributions of the starting values $S_1, ..., S_N$ are not supposed to depend on the parameters of model (2). Note that assuming ψ_i to be Gaussian does not limit the support of θ_i within the stability region, and thus even allows for explosive individual trajectories with a positive probability. This is not necessarily a drawback, but rather reflects the flexibility of model (1).

2. Estimation of the model

The likelihood function defined by Y_i conditionally on S_i and ψ_i takes the form

$$L_{Y_i|S_i,\psi_i}(\theta, \sigma^2) \propto \sigma^{-(n_i - p)} \exp\{-\frac{1}{2\sigma^2}(Y_i - X_i^*\theta - Z_i\psi_i)'(Y_i - X_i^*\theta - Z_i\psi_i)\}.$$

By taking expectations with respect to ψ_i we get the actual likelihood conditional only on S_i. After some manipulations, the resulting loglikelihood can be written as

(3) $$\log\ L_{Y_i|S_i}(\theta, \Sigma_o, \sigma^2) \simeq -\frac{1}{2}\log\ \mid V_i \mid -\frac{1}{2}(Y_i - X_i^*\theta)' V_i^{-1}(Y_i - X_i^*\theta)$$

where

$$V_i = \sigma^2\{I - Z_i(\Sigma_o^{-1} + Z_i'Z_i)^{-1}Z_i'\}^{-1} = \sigma^2(I + Z_i\Sigma_o Z_i') \quad .$$

The result resembles the likelihood corresponding to a standard mixed linear model (cf. Longford, 1993, pp. 100-106), but its interpretation is now different from the regular case, because the y_{it} observations appear in the X_i^* and Z_i matrices as well. Anyway, the astonishingly simple form of (3) makes likelihood-based inference methods very attractive. The sampling distributions of the ML estimators and some testing problems have been discussed by Rahiala (1999).

Conventional computational procedures can also be used to estimate the random effects, because it can be shown that

(4) $$E(\psi_i \mid Y_i, X_i^*, S_i) = (\Sigma_o^{-1} + Z_i'Z_i)^{-1}Z_i'(Y_i - X_i^*\theta) = \sigma^2\Sigma_o Z_i' V_i^{-1}(Y_i - X_i^*\theta) \quad .$$

REFERENCES

Arellano, M. (1989). An efficient GLS estimator of triangular models with covariance restrictions. Journal of Econometrics 42, 267-273.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58, 277-297

Holtz-Eakin, D. , Newey, W. and Rosen, H. S. (1988). Estimating vector autoregressions with panel data. Econometrica 56, 1371-1395

Longford, N. T. (1995). Random Coefficient Models. Oxford University Press. Oxford.

Rahiala, M. (1999). Random coefficient autoregressive models for longitudinal data. Forthcoming in Biometrika.

Rosner, B. , Munoz, A. , Tager, I. , Speizer, F. and Weiss, S. (1985). The use of an autoregressive model for the analysis of longitudinal data in epidemiologic studies. Statistics in Medicine 4, 457-467.

RÉSUMÉ

Nous étudions l'utilisation des modèles autorégressifs linéaires avec des coefficients aléatoires dans l'analyse des données longitudinales. La fonction de vraisemblance et les méthodes d'inférence basées sur cela se montrent étonnamment simples.