MATHEMATICAL MODELING AND COMPUTING, Vol. 10, No. 3, pp. 682–692 (2023)

Estimation in short-panel data models with bilinear errors

Lmakri A.[1], Akharif A.[2], Mellouk A.[3]

[1] *National Higher School of Arts and Crafts (ENSAM), Hassan II University, Casablanca, Morocco*
[2] *Mathematics and Applications Laboratory, FSTT, Abdelmalek Essaadi University Tetouan, Morocco*
[3] *Regional Center for Education and Training Trades, Tangier, Morocco*

(Received 16 February 2023; Revised 12 July 2023; Accepted 13 July 2023)

Many estimation methods have been proposed for the parameters of the regression models with serially correlated errors. In this work, we develop an asymptotic theory for estimation in the short panel data models with bilinear error. We propose a comparative study by simulation between several estimators (adaptive, ordinary and weighted least squares) for the coefficients of panel data models when the errors are bilinear serially correlated. As a consequence of the uniform local asymptotic normality property, we obtain adaptive estimates of the parameters. Finally, we illustrate the performance of the proposed estimators via Monte Carlo simulation study. We show that the adaptive estimates are more efficient than the weighted and ordinary least squares estimates.

Keywords: *adaptive estimate; bilinear models; panel regression models; weighted least squares; ordinary least squares.*

2010 MSC: 62F03, 62F05, 62G10, 62G20 **DOI:** 10.23939/mmc2023.03.682

1. Introduction

Regression models have many applications in the areas of economics, accounting, finance, engineering, environmental studies, medical sciences, etc. One of the tools to analyze large and high-dimensional data sets is the panel data model. In this last, individuals (persons, firms, cities, etc.) are observed at several points in time (days, years, before and after treatment, etc.). Many researchers (in these disciplines) in recent years have relied on panel data to model the behavior of individuals. They have done so because panel data models allow them to control for temporally persistent unobserved differences among individuals that in many instances may bias estimates obtained from cross sections.

In classical panel data models, there is an important assumption, which is the error distribution is normal i.d., with mean zero and finite variance. This standard assumption of the normality for the error term in regression models in many situations do not agree with real data sets. In order to take into account, the first-order serial correlation in the remaining perturbations, [1] introduced the error component model for the random-effects model. Furthermore, the fixed-effects model had extended by [2]. Both studies considered the autoregressive-AR(1) specification on the remainder disturbances. However, the moving average-MA(1) model is a viable alternative (see [3]). Moreover, [4] had considered the stochastic parameter panel data models when the errors are first-order serially correlated and he had examined different estimators for these models. Not only but again various manuscripts treat the problem of correlated errors in regression models in which the errors follow the linear models such as autoregressive (AR), moving average (MA) (e.g. [5]), the mixed autoregressive and moving average (ARMA) models (e.g. [6]), or the nonlinear models such as the random coefficient autoregressive (RCAR) model, the autoregressive conditional heteroscedasticity (ARCH) model, fractional-ARIMA and bilinear models (e.g. [7–10], etc.).

This paper focuses on short panels with relatively few time periods and many individuals (small T and large n). We consider the general situation where the error distribution is not necessarily normal and bilinear serially correlated. The model is defined, for $i = 1, 2, \dots, n$ and $t = 1, 2, \dots, T$, by:

$$y_{i,t} = \mu + \beta' x_{i,t} + e_{i,t}, \tag{1}$$

where $y_{i,t}$ is the observation on the i^{th} cross-sectional unit for the t^{th} time period, $x_{i,t}$ denotes the vector of observations on the non-stochastic regressors, and $(\mu, \beta')' \in \mathbb{R}^{K+1}$ is the corresponding regression coefficients, which is the unknown parameters of interest. The additive error terms $e_{i,t}$ are following a simple case of panel bilinear model, which takes the following form:

$$e_{i,t} = b e_{i,t-l} \varepsilon_{i,t-k} + \varepsilon_{i,t} \quad \text{with} \quad l > k \geqslant 1, \tag{2}$$

here $\varepsilon_{i,t}$ is a white noise with mean zero, finite variance σ^2 for all i and t, and density $f\colon \varepsilon \mapsto f(\varepsilon) := (1/\sigma) f_1(\varepsilon/\sigma)$, where f_1 belongs to the adequate class of standardized densities

$$\mathcal{F}_0 := \left\{ f_1\colon f_1(u) > 0 \ \forall u \in \mathbb{R}, \int_{-1}^{1} f_1(u)\, du = 0.5 = \int_{-\infty}^{0} f_1(u)\, du \right\}.$$

Probabilistic properties such as stationarity and invertibility have been studied in [11] remains valid in panel bilinear model (2).

Denote by $\mathcal{F}_{i,t}(\varepsilon)$ and $\mathcal{F}_{i,t}(e)$ the σ-algebras generated by $\{\varepsilon_{i,s} | s \leqslant t\}$ and $\{e_{i,s} | s \leqslant t\}$, respectively. Then,

- Equation (2) admits a unique stationary solution $e_{i,t}$ if and only if $b^2\sigma^2 < 1$, and given by

$$e_{i,t} = \varepsilon_{i,t} + \sum_{j=1}^{\infty} b^j \varepsilon_{i,t-lj} \prod_{s=1}^{j} \varepsilon_{i,t-k-(s-1)l}.$$

- Equation (2) is invertible if and only if $2b^2\sigma^2 < 1$, in this case, one can write

$$\varepsilon_{i,t} = e_{i,t} + \sum_{j=1}^{\infty} (-b)^j e_{i,t-kj} \prod_{s=1}^{j} e_{i,t-l-(s-1)k}.$$

Let $y = (y_{1,1}, \ldots, y_{n,T})'_{nT\times 1}$, $e = (e_{1,1}, \ldots, e_{n,T})'_{nT\times 1}$, $\beta = (\beta_0 = \mu, \beta_1, \ldots, \beta_K)'_{(K+1)\times 1}$, and

$$X = \begin{pmatrix} 1 & x_{1;1,1} & x_{2;1,1} & \cdots & x_{K;1,1} \\ 1 & x_{1;1,2} & x_{2;1,2} & \cdots & x_{K;1,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1;n,T} & x_{2;n,T} & \cdots & x_{K;n,T} \end{pmatrix}_{nT\times(K+1)}.$$

The matrix form of model (1)–(2) yield:

$$y = X\beta + e.$$

In the literature, several estimation methods have appeared very successful, where parameters of the model considered are used for the inference results concerning different subjects of interest. In this paper, we show that the adaptive estimation method has an efficiency gain and asymptotic performance.

This paper is organized as follows. Section 2 provides three estimation methods. Subsection 2.1 offers the adaptive estimators for parameters model with correlated errors. The feasible versions of ordinary least squares (OLS) and weighted least squares (WLS) estimators have been suggested in subsection 2.2. Section 3 contains the Monte Carlo simulation study for efficiency comparisons of different estimators. Finally, concluding remarks are included in Section 4.

2. Estimation methods

The most popular methods for fitting multiple regression models among others are the maximum likelihood and weighted least squares (see [12]). These methods have compared to other methods in many studies. Moreover, many papers have claimed that maximum likelihood can be efficient (more or less) than weighted least squares or ordinary least squares.

In 1981, Hausman and Wise claimed in their Section 10.3 that the maximum likelihood estimates are more efficient. They state that the gain in efficiency from using maximum likelihood instead of weighted least squares is small in some cases and the relative efficiency of maximum likelihood becomes substantial in other cases. On the other hand, many papers suggest that maximum likelihood is superior or inferior to weighted and least squares and other approaches.

The uniform local asymptotic normality (ULAN) of model (1)–(2) is established in [13] by using the quadratic mean differentiability of the density characterizing the model. Once the ULAN property is proved, the limit distribution and the asymptotic optimality of the locally asymptotically minimax (LAM) estimators are described in this latter. Thereafter, we look into the limit distributions of both weighted and ordinary least squares estimators of the model parameters.

2.1. Adaptive estimation

Model (2) can be looked as one-type of ARCH model. [14] constructed adaptive and hence efficient estimators in a general GARCH. Their analysis is based on a general LAN theorem for time series models. Furthermore, [15] considered adaptive estimation in nonstationary ARMA models with the noise sequence satisfying a generalized ARCH process.

The ULAN property is an important notion under which we can define a notion of asymptotic efficiency for estimators (see [16]). Before starting the results, we remind for the reader's convenience some notations and main assumptions under ULAN statistical experiments.

Consider a sequence of random variables $\big[y_1^{(n)'}, y_2^{(n)'}, \dots, y_n^{(n)'}\big]'$ with $y_i^{(n)} := (y_{i,1}, \dots, y_{i,T})'$ is an observed series of length nT. Let $\mathbf{P}^{(n)}_{\theta;f_1}$ the hypothesis under which $y_i^{(n)}$ is generated by the model (1)–(2) and $\theta := (\mu, \beta', \sigma^2, b) \in \mathbb{R}^{K+3}$ is the parameter to be estimated in the sequel.

As usual, it is assumed that the vector of initial values

$$e_0^{(n)} := \Big\{(e^{(n)}_{i,1-l}\varepsilon_{i,1-k}, e^{(n)}_{i,2-l}\varepsilon_{i,2-k}, \dots, e^{(n)}_{i,k-l}\varepsilon_{i,0}, e^{(n)}_{i,k+1-l}, \dots, e^{(n)}_{i,-1}, e^{(n)}_{i,0}), i = 1, \dots, n\Big\}$$

is observable for each individual i.

Some technical assumptions about the innovation density f_1 are required to establishing ULAN property:

(A.1) $f_1 \in \mathcal{F}_0$;

(A.2) f_1 is $\mathcal{C}^1$ on $\mathbb{R}$, with the first derivative f_1' and letting $\Phi_{f_1} = -f_1'/f_1$.

Assume that

$$I(f_1) := \int_{\mathbb{R}} \Phi^2_{f_1}(u) f_1(u)\, du, \quad J(f_1) := \int_{\mathbb{R}} u^2 \Phi^2_{f_1}(u) f_1(u)\, du, \quad \text{and} \quad K(f_1) := \int_{\mathbb{R}} u \Phi^2_{f_1}(u) f_1(u)\, du$$

are finite.

Let $\tau^{(n)} := \big(\tau_1^{(n)}, \tau_2^{(n)'}, \tau_3^{(n)}, \tau_4^{(n)}\big)'$ be a sequence of real vectors in $\mathbb{R}^{K+3}$ such that $\tau^{(n)'}\tau^{(n)}$ is uniformly bounded as $n \to \infty$. In addition, we consider a $K \times K$ nonsingular matrix $\mathbf{K}^{(n)}$ defined as: $\mathbf{K}^{(n)} := (\mathbf{C}^{(n)})^{-1/2}$ with $\mathbf{C}^{(n)} := \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T x_{i,t}x'_{i,t}$. Letting

$$\nu^{(n)} := n^{-1/2} \begin{bmatrix} 1 & \mathbf{0} & 0 & 0 \\ \mathbf{0} & \mathbf{K}^{(n)} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

with these notations, $\mathbf{P}^{(n)}_{\theta+\nu^{(n)}\tau^{(n)};f_1}$ consist of a sequence of hypotheses approaching $\mathbf{P}^{(n)}_{\theta;f_1}$ (in the sense of contiguity).

Define the standardized residuals for $i = 1, 2, \dots, n$ and $t = 1, 2, \dots, T$ as:

$$Z_{i,t} = Z_{i,t}(\theta) := \sigma^{-1}(y_{i,t} - \mu - \beta' x_{i,t}) = \sigma^{-1} e_{i,t}.$$

Clearly, under $\mathbf{P}^{(n)}_{\theta;f_1}$, $Z_{i,t}$ is a nT-tuple of random variables with probability density f_1.

By virtue of ULAN property, the class of maximum likelihood estimators of θ can be constructed using a $(\nu^{(n)})^{-1}$-consistent estimator of θ as a preliminary estimator. In our model, using a preliminary $(\nu^{(n)})^{-1}$-consistent estimator $\tilde{\theta}_n$ of θ, i.e. a sequence of statistics such that:

(i) $(\nu^{(n)})^{-1}(\tilde{\theta}_n - \theta) = O_P(1)$, under $\mathbf{P}^{(n)}_{\theta;f_1}$, as $n \to \infty$.

(ii) For all $c > 0$ fixed, the number of possible values of $\tilde{\theta}_n$ in $\mathcal{B} = \big\{u \in \mathbb{R}^{K+3} \colon \, \| (\nu^{(n)})^{-1}(u - \theta) \| \leqslant c\big\}$ is bounded as $n \to \infty$.

From ULAN property, we can construct sequences of estimates which are locally asymptotically minimax. These LAM estimators of θ can be defined as:

$$\widehat{\theta}_n := \tilde{\theta}_n + \nu^{(n)}\Gamma_{f_1}^{-1}(\tilde{\theta}_n)\Delta_{f_1}^{(n)}(\tilde{\theta}_n); \tag{3}$$

with $(K+3)$-dimensional central sequence

$$\Delta_{f_1}^{(n)}(\theta) := \begin{bmatrix} \Delta_{f_1;1}^{(n)}(\theta) \\ \Delta_{f_1;2}^{(n)}(\theta) \\ \Delta_{f_1;3}^{(n)}(\theta) \\ \Delta_{f_1;4}^{(n)}(\theta) \end{bmatrix} := \begin{bmatrix} \frac{n^{-1/2}}{\sigma}\sum_{i=1}^{n}\sum_{t=1}^{T}\Phi_{f_1}(\frac{\varepsilon_{i,t}}{\sigma}) \\ \frac{n^{-1/2}}{\sigma}\sum_{i=1}^{n}\sum_{t=1}^{T}\Phi_{f_1}(\frac{\varepsilon_{i,t}}{\sigma})(\mathbf{K}^{(n)})' x_{i,t} \\ \frac{n^{-1/2}}{2\sigma^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\Phi_{f_1}(\frac{\varepsilon_{i,t}}{\sigma})Z_{i,t} - 1\right] \\ \frac{n^{-1/2}}{\sigma}\sum_{i=1}^{n}\sum_{t=l+1}^{T}\Phi_{f_1}(\frac{\varepsilon_{i,t}}{\sigma})\sum_{j=1}^{\infty} j(-b)^{j-1} e_{i,t-k}\prod_{s=1}^{j} e_{i,t-l-(s-1)k} \end{bmatrix},$$

and $(K+3)\times(K+3)$-information matrix

$$\Gamma_{f_1}(\theta) := \begin{bmatrix} \frac{TI(f_1)}{\sigma^2} & 0 & \frac{TK(f_1)}{2\sigma^3} & 0 \\ \mathbf{0} & \frac{TI(f_1)\mathbf{I}_K}{\sigma^2} & \mathbf{0} & \mathbf{0} \\ \frac{TK(f_1)}{2\sigma^3} & 0 & \frac{T}{4\sigma^4}\left(\frac{J(f_1)}{1-b^2\sigma^2} - 1\right) & 0 \\ 0 & 0 & 0 & \Gamma_{f_1;44}(\theta) \end{bmatrix}.$$

In current terminology, following [17], an estimator sequence $\widehat{\theta}_n$ with such a property is called *best regular*. It is also *asymptotically efficient* in the sense of [18]. The following theorem presents the asymptotic properties of the estimator proposed above.

Theorem 1. *Assume that (A.1) and (A.2) hold. Then, under* $\mathbf{P}_{\theta;f_1}^{(n)}$,

$$(\nu^{(n)})^{-1}(\widehat{\theta}_n - \theta) = \Gamma_{f_1}^{-1}(\theta)\Delta_{f_1}^{(n)}(\theta) + o_P(1), \quad \text{as} \quad n \to \infty, \tag{4}$$

and the asymptotic distribution of $(\nu^{(n)})^{-1}(\widehat{\theta}_n - \theta)$ *is:*

$$(\nu^{(n)})^{-1}(\widehat{\theta}_n - \theta) \xrightarrow[n\to\infty]{\mathcal{D}} \begin{cases} \mathcal{N}\big(0, \Gamma_{f_1}^{-1}(\theta)\big), & \text{under } \mathbf{P}_{\theta;f_1}^{(n)}; \\ \mathcal{N}\big(\tau, \Gamma_{f_1}^{-1}(\theta)\big), & \text{under } \mathbf{P}_{\theta+\nu^{(n)}\tau^{(n)};f_1}^{(n)}. \end{cases} \tag{5}$$

Proof. The first equality (4) follows from the asymptotic linearity of the central sequence $\Delta_{f_1}^{(n)}(\theta)$, the discreteness of $\tilde{\theta}_n$, and the fact that $\Gamma_{f_1}(\tilde{\theta}_n)$ is consistent for $\Gamma_{f_1}(\theta)$. Furthermore, the asymptotic normality of $\widehat{\theta}_n$ in (3) under various optimality criteria, can be deduced from the limite distribution of $\Delta_{f_1}^{(n)}(\theta)$ which converges in distribution to a $(K+3)^2$-variate normal distribution with mean zero under $\mathbf{P}_{\theta;f_1}^{(n)}$, mean $\Gamma_{f_1}(\theta)\tau$ under $\mathbf{P}_{\theta+\nu^{(n)}\tau^{(n)};f_1}^{(n)}$ and covariance matrix $\Gamma_{f_1}(\theta)$ under both (cf. [13]). ■

Note that we have, in fact, a class of asymptotically optimal estimators since any $(\nu^{(n)})^{-1}$-consistent estimator $\tilde{\theta}_n$ yields a locally asymptotically minimax estimator. Finally, it should be mentioned that $\widehat{\theta}_n$ is known to be asymptotically optimal under various other optimality criteria.

Adaptivity by estimating the score function. Herein we estimate the score function $\Phi_{f_1} = -f_1'/f_1$ by using the kernel estimator method (see [20]).

Let a_n and b_n are two real sequences of positive terms converging to zero. The estimator of the score function Φ_{f_1} associated with the residual series $Z^{(n)} = \left(Z_{1,1}^{(n)}(\theta), \dots, Z_{n,T}^{(n)}(\theta)\right)'$ is defined by:

$$\widehat{\Phi}^{(n)}(x) = -\frac{f_1^{(n)'}(x, Z^{(n)})}{b_n + f_1^{(n)}(x, Z^{(n)})}, \tag{6}$$

with

$$f_1^{(n)}(x, Z^{(n)}) = \frac{1}{nTa_n} \sum_{i=1}^{n} \sum_{t=1}^{T} k\left(\frac{x - Z_{i,t}^{(n)}(\theta)}{a_n}\right)$$

and

$$f_1^{(n)'}(x, Z^{(n)}) = \frac{1}{nTa_n^2} \sum_{i=1}^{n} \sum_{t=1}^{T} k'\left(\frac{x - Z_{i,t}^{(n)}(\theta)}{a_n}\right), \quad \text{for} \quad x \in \mathbb{R},$$

where the kernel k satisfies Condition K of [19].

In order to construct an adaptive estimates, our approach consists of replacing the score function Φ_{f_1} in the central sequence $\Delta_{f_1}^{(n)}(\theta)$ by an estimator $\widehat{\Phi}^{(n)}$. Now consider the adaptive central sequence $\widehat{\Delta}^{(n)}(\cdot)$ defined similarly as in (3).

Corollary 1. *Let $\tilde{\theta}_n$ be a $(\nu^{(n)})^{-1}$-consistent estimator of θ. Then, under $\mathbf{P}_{\theta;f_1}^{(n)}$ and as $n \to \infty$ we have:*

$$\begin{cases} \widehat{\Delta}^{(n)}(\tilde{\theta}_n) - \Delta_{f_1}^{(n)}(\tilde{\theta}_n) = o_P(1), \\ \widehat{\Gamma}(\tilde{\theta}_n) - \Gamma_{f_1}(\tilde{\theta}_n) = o_P(1). \end{cases}$$

Then, the adaptive estimate is presented by

$$\widehat{\widehat{\theta}}_n = \tilde{\theta}_n + \nu^{(n)} \widehat{\Gamma}^{-1}(\tilde{\theta}_n) \widehat{\Delta}^{(n)}(\tilde{\theta}_n). \tag{7}$$

Under $\mathbf{P}_{\theta;f_1}^{(n)}$ and as $n \to \infty$, we have:

$$(\nu^{(n)})^{-1}(\widehat{\widehat{\theta}}_n - \theta) = \Gamma_{f_1}^{-1}(\theta) \Delta_{f_1}^{(n)}(\theta) + o_P(1), \tag{8}$$

and their asymptotic normality is established along the same lines of (5).

2.2. Ordinary and weighted least squares estimators

Regression modeling is one of the important statistical techniques used in many applications. The classical theory shows that the OLS estimator of the unknown vector of regression coefficients is the best linear unbiased estimator and asymptotically efficient when the errors have the same variance (homoscedastic). But in many practical problems, the OLS assumption of constant variance is violated in the errors. In other words, the error variances are unequal (heteroscedastic) and therefore the optimal properties of the OLS are lost. If the variances are known, the best estimator, in this case, is the WLS estimator with the reciprocals of the variances as the weights. However, usually, the error variances are unknown. A natural and frequently used approach is to obtain estimates of the error variances from the observed data and use the reciprocals of the variance estimates as the weights in the WLS.

In this subsection, OLS and WLS estimators of the coefficients of a multiple regression model in panel data with serially correlated errors are derived and their limit distributions are also studied.

We consider the problem of estimating β in a multiple regression model with heteroscedastic error variances:

$$y = X\beta + e, \tag{9}$$

where y is an nT-vector of observations $y_{i,t}$ $(i = 1, 2, \dots, n;\ t = 1, 2, \dots, T)$, X is an $nT \times (K+1)$ full rank matrix of known constants $(nT > K+1)$, β is an $(K+1)$-vector of regression parameters and e is an nT-vector of random variables $e_{i,t}$ with mean zero and conditional dispersion matrix (non constant variance-covariance matrix)

$$V = \operatorname{diag}\left(\sigma_{1,1}^2, \dots, \sigma_{n,T}^2\right),$$

where $\sigma_{i,t}^2 = \sigma^2 + b^2\sigma^2 e_{i,t-l}^2$ is the unknown error variance associated with $e_{i,t}$.

2.2.1. Ordinary least squares estimator

When we use ordinary least squares to estimate linear regression, we minimize the mean squared error:

$$\mathrm{MSE}(\beta) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{y=1}^{T}(y_{i,t} - x'_{i,t}\beta)^2.$$

The solution, namely the ordinary least squares estimator of β, and given in the vector form:

$$\widehat{\beta}_{OLS} = (X'X)^{-1}X'y. \tag{10}$$

2.2.2. Weighted least squares estimator

Herein we construct a weighted least squares estimator for the regression parameters β based on the estimated error parameters b and σ^2. Furthermore, we show that it is asymptotically more efficient than the ordinary least squares estimator.

Since $\mathrm{Var}(y_{i,t}/e_{i,t-l}) = \sigma^2_{i,t}$ depends on $e_{i,t-l}$ one may consider a conditional WLS estimator of β in order to improve the efficiency. Overall, the WLS are the popular method of solving the problem of heteroscedasticity in regression models. In this paragraph, we derive such an estimator and study its properties.

We could instead minimize the *weighted* mean squared error,

$$\mathrm{WMSE}(\beta, w_{1,1}, \ldots, w_{n,T}) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{i,t}(y_{i,t} - x'_{i,t}\beta)^2,$$

where $w_{i,t} = \frac{1}{\sigma^2_{i,t}}$. As a special case where all the weights $w_{i,t} = 1$, we see that this coincides with OLS estimator. By the same kind of linear algebra we used to solve the ordinary linear least squares problem, we can solve the above problem. Let W be the matrix with the $w_{i,t}$ on the diagonal and zeroes everywhere else, then

$$\mathrm{WMSE} = \frac{1}{nT}(y - X\beta)'W(y - X\beta).$$

Differentiating with respect to β, we get as the gradient

$$\nabla_\beta \mathrm{WMSE} = \frac{2}{nT}(-X'Wy + X'WX\beta).$$

Setting this to zero at the optimum and solving, the conditional weighted least squares estimator of β is given by:

$$\widehat{\beta}_{WLS} = (X'W_{\widehat{\eta}}X)^{-1}X'W_{\widehat{\eta}}y, \tag{11}$$

where

$$W_\eta = V^{-1} = \mathrm{diag}\left(\frac{1}{\sigma^2 + b^2\sigma^2 e^2_{1,-2}}, \ldots, \frac{1}{\sigma^2 + b^2\sigma^2 e^2_{n,T-l}}\right)$$

and $\widehat{\eta}$ is an estimator of $\eta := (\sigma^2, b^2\sigma^2)'$. The following paragraph is concerned with the problem of estimating η.

Let $R_{i,t}(\beta) = y_{i,t} - E(y_{i,t}/e_{i,t-l})$ and denote $\sigma^2_{i,t} = \sigma^2_{i,t}(\eta)$ defined as:

$$\sigma^2_{i,t}(\eta) = E\big(R^2_{i,t}(\beta)/e_{i,t-l}\big) = \sigma^2 + b^2\sigma^2 e^2_{i,t-l}.$$

A conditional least squares estimator $\widehat{\eta}(\beta)$ of η can be obtained by minimizing

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\big(R^2_{i,t}(\beta) - \sigma^2_{i,t}(\eta)\big)^2.$$

Thus, $\widehat{\eta}(\beta)$ is given by a solution of the equation:

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\big(R^2_{i,t}(\beta) - \sigma^2_{i,t}(\eta)\big)\frac{\partial\sigma^2_{i,t}(\eta)}{\partial\eta} = 0.$$

The estimate $\widehat{\eta}(\beta)$ when β is known it is then seen to be:

$$\widehat{\eta}(\beta) = (U'U)^{-1}U'V(\beta), \tag{12}$$

where

$$U = \begin{pmatrix} 1 & e_{1,1-l}^2 \\ \vdots & \vdots \\ 1 & e_{n,T-l}^2 \end{pmatrix} \quad \text{and} \quad V(\beta) = \big(R_{1,1}^2(\beta), \ldots, R_{n,T}^2(\beta)\big)'.$$

In order to derive the limit distribution of (12), we suppose that $E(|e_{i,t}|^8) < \infty$, a sufficient condition for this is that $E(|\varepsilon_{i,t}|^8) < \infty$. So, the limit distribution of $\widehat{\eta}$ is given by the following theorem.

Theorem 2. *Let $\widehat{\eta} = \widehat{\eta}(\widehat{\beta})$, where $\widehat{\beta}$ is given by (10), we have*

$$\sqrt{n}(\widehat{\eta} - \eta) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\big(\mathbf{0}, K^{-1}LK^{-1}\big), \tag{13}$$

where

$$K = E\left[\left(\frac{\partial \sigma_{i,t}^2(\eta)}{\partial \eta}\right)\left(\frac{\partial \sigma_{i,t}^2(\eta)}{\partial \eta}\right)'\right] = \begin{bmatrix} 1 & E(e_{i,t}^2) \\ E(e_{i,t}^2) & E(e_{i,t}^4) \end{bmatrix},$$

and

$$L = E\left[\left(R_{i,t}^2(\beta) - \sigma_{i,t}^2(\eta)\right)^2 \left(\frac{\partial \sigma_{i,t}^2(\eta)}{\partial \eta}\right)\left(\frac{\partial \sigma_{i,t}^2(\eta)}{\partial \eta}\right)'\right] = \begin{bmatrix} L_{11} & L_{12} \\ L_{12} & L_{22} \end{bmatrix},$$

with

$$\begin{aligned} L_{11} &= \big(1 - b^4\sigma^4\big)E(e_{i,t}^4) - 2\sigma^2 E(e_{i,t}^2) + \sigma^4, \\ L_{12} &= b^4\big(E(\varepsilon_{i,t}^4) - \sigma^4\big)E(e_{i,t}^6) + 4b^2\sigma^4 E(e_{i,t}^4) + \big(E(\varepsilon_{i,t}^4) - \sigma^4\big)E(e_{i,t}^2), \\ L_{22} &= b^4\big(E(\varepsilon_{i,t}^4) - \sigma^4\big)E(e_{i,t}^8) + 4b^2\sigma^4 E(e_{i,t}^6) + (E(\varepsilon_{i,t}^4) - \sigma^4)E(e_{i,t}^4), \end{aligned}$$

and

$$\begin{aligned} E(e_{i,t}^2) &= \frac{\sigma^2}{1 - b^2\sigma^2}, \\ E(e_{i,t}^4) &= \frac{6b^2\sigma^4 E(e_{i,t}^2) + E(\varepsilon_{i,t}^4)}{1 - b^4 E(\varepsilon_{i,t}^4)}, \\ E(e_{i,t}^6) &= \frac{1}{1 - b^6 E(\varepsilon_{i,t}^6)}\big(15b^4\sigma^2 E(e_{i,t}^4)E(\varepsilon_{i,t}^4) + 15b^2\sigma^2 E(e_{i,t}^2)E(\varepsilon_{i,t}^4) + E(\varepsilon_{i,t}^6)\big), \\ E(e_{i,t}^8) &= \frac{1}{1 - b^8 E(\varepsilon_{i,t}^8)}\big(28b^6\sigma^2 E(e_{i,t}^6)E(\varepsilon_{i,t}^6) + 70b^4 E(e_{i,t}^4)(E(\varepsilon_{i,t}^4))^2 + 28b^2\sigma^2 E(e_{i,t}^2)E(\varepsilon_{i,t}^6)\big). \end{aligned}$$

Proof. The limit distribution is established using the following lemma.

Lemma 1. *We have as $n \to \infty$ and T fixed:*

(i) $n^{-1}U'U \xrightarrow[n\to\infty]{a.s} TK$,

(ii) $n^{-1}U'\big(V(\beta) - U\eta\big) \xrightarrow[n\to\infty]{a.s} 0$,

(iii) $n^{-1/2}U'\big(V(\widehat{\beta}) - V(\beta)\big) \xrightarrow[n\to\infty]{\mathcal{P}} 0$,

(iv) $n^{-1/2}U'\big(V(\beta) - U\eta\big) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, T^2L)$.

Then, we consider the decomposition

$$\sqrt{n}\big(\widehat{\eta} - \eta\big) = \sqrt{n}\big(\widehat{\eta} - \widehat{\eta}(\beta)\big) + \sqrt{n}\big(\widehat{\eta}(\beta) - \eta\big).$$

First, note that:

$$\begin{aligned} \sqrt{n}(\widehat{\eta} - \widehat{\eta}(\beta)) &= \sqrt{n}(U'U)^{-1}U'(V(\widehat{\beta}) - V(\beta)) \\ &= (TK)^{-1}n^{-1/2}U'\big(V(\widehat{\beta}) - V(\beta)\big) + o_p(1) \\ &= o_p(1), \end{aligned}$$

by (i) and (iii) of the above lemma. Moreover,

$$\begin{aligned}\sqrt{n}(\widehat{\eta}(\beta)-\eta) &= \sqrt{n}(U'U)^{-1}U'\big(V(\beta)-U\eta\big)\\ &= (TK)^{-1}n^{-1/2}U'\big(V(\beta)-U\eta\big)+o_p(1)\\ &\xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\big(\mathbf{0},K^{-1}LK^{-1}\big),\end{aligned}$$

by (i) and (iv) of the above. Then, the theorem is proved. ■

Finally, the theorem below shows the asymptotic normality of $\widehat{\beta}_{WLS}$.

Theorem 3. *Denote $\widehat{\beta}_{WLS}=\widehat{\beta}_{WLS}(\widehat{\eta})$, as $n\to\infty$ we have*

$$\sqrt{n}(\widehat{\beta}_{WLS}-\beta)\xrightarrow[n\to\infty]{\mathcal{D}}\mathcal{N}\big(\mathbf{0},(X'W_\eta X)^{-1}\big). \tag{14}$$

Stationary errors with known autocovariances are presented among all un-weighted linear estimators, including OLS estimators, there are WLS estimators that have minimal variance. The estimator $\widehat{\beta}_{WLS}$ is obtained in two steps. The first step is to obtain an estimator $W_{\widehat{\eta}}$. The second step is the calculation outlined in (11).

3. Monte Carlo simulation study

In this section, we will make Monte Carlo simulation study to compare the performance of three estimators proposed of the regression parameters in sampling experiments.

Let $y=(y_{1,1},\ldots,y_{n,T})$ be a $nT\times 1$ vector of dependent variable observations. We set $K=2$, let $\beta=(\beta_0=\mu,\beta_1,\beta_2)$ be the 3×1 vector of regression parameters, and $e=(e_{1,1},\ldots,e_{n,T})$ be a $nT\times 1$ vector of additive errors. Consider the model:

$$\begin{cases} y_{i,t}=\mu+\beta_1 x_{1;i,t}+\beta_2 x_{2;i,t}+e_{i,t},\\ e_{i,t}=b e_{i,t-3}\varepsilon_{i,t-1}+\varepsilon_{i,t}.\end{cases} \tag{15}$$

We take:

- ⋆ $n=30,50,70,100$ and $T=15$;
- ⋆ $\beta=(0.5,1,-1)'$;
- ⋆ $x_{i,t}=\begin{bmatrix}x_{1;i,t}\\ x_{2;i,t}\end{bmatrix}\sim\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}5 & 2\\ 2 & 10\end{bmatrix}\right)$;
- ⋆ $b=(-0.5,-0.3,-0.1,0,0.1,0.3,0.5)$;
- ⋆ the $\varepsilon_{i,t}$'s are i.i.d. $\mathcal{N}(0,0.3)$.

Define

$$X=\begin{pmatrix}1 & x_{1;1,1} & x_{2;1,1}\\ 1 & x_{1;1,2} & x_{2;1,2}\\ \vdots & \vdots & \vdots\\ 1 & x_{1;n,T} & x_{2;n,T}\end{pmatrix}_{(nT\times 3)}.$$

It is also convenient to write the linear model (15) in vector-matrix notation: $y=X\beta+e$.

Thus,

– The ordinary least squares estimator of β is: $\widehat{\beta}_{OLS}=(X'X)^{-1}X'y$.

– The weighted least squares estimator of β is given by: $\widehat{\beta}_{WLS}=(X'W_{\widehat{\eta}}X)^{-1}X'W_{\widehat{\eta}}y$.

– The adaptive estimator of β is defined as: $\widehat{\widehat{\beta}}_n=\tilde{\beta}_n+n^{-1/2}\begin{bmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{K}^{(n)}\end{bmatrix}_{3\times 3}\widehat{\Gamma}^{-1}(\tilde{\beta}_n)\widehat{\Delta}^{(n)}(\tilde{\beta}_n)$.

In defining $\widehat{\widehat{\beta}}_n$, we require the existence of consistent estimators $\tilde{\beta}_n$, $\widehat{I}_n$ and $\widehat{\sigma}_n^2$ of β, $I(f_1)$ and σ^2, respectively. For this purpose one can use:

$$\tilde{\beta}_n=\widehat{\beta}_{OLS},\quad \widehat{I}_n=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\big[\widehat{\Phi}^{(n)}\big(\varepsilon_{i,t}(\tilde{\theta}_n)\big)\big]^2,\quad\text{and}\quad \widehat{\sigma}_n^2=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\big[\varepsilon_{i,t}\big(\tilde{\theta}_n\big)\big]^2.$$

The sequences $a_n = b_n = n^{-1/16}$ were adopted with the kernel $k(x) = \frac{c}{(1+x^6)}$, with $c = 1/\int_{\mathbb{R}}(1+x^6)^{-1}dx \approx 2.094$. Then, the adaptive central sequence and its information matrix are respectively;

$$\widehat{\Delta}^{(n)}(\widehat{\theta}) := \frac{n^{-1/2}}{\widehat{\sigma}_n} \begin{pmatrix} \sum_{i=1}^{n}\sum_{t=1}^{T} \widehat{\Phi}^{(n)}\left(\frac{\varepsilon_{i,t}}{\widehat{\sigma}_n}\right) \\ \sum_{i=1}^{n}\sum_{t=1}^{T} \widehat{\Phi}^{(n)}\left(\frac{\varepsilon_{i,t}}{\widehat{\sigma}_n}\right)(\mathbf{K}^{(n)})' x_{i,t} \end{pmatrix}_{3\times 1} \quad \text{and} \quad \widehat{\Gamma}(\widehat{\theta}) := \frac{T}{\widehat{\sigma}_n^2}\widehat{I}_n \mathbf{I}_3.$$

In the following, we compare the performance of three estimators $\widehat{\beta}_{OLS}$, $\widehat{\beta}_{WLS}$ and $\widehat{\widehat{\beta}}_n$ of $\beta = (\beta_0, \beta_1, \beta_2)'$.

Table 1. Ratio of mean-squared errors of $\widehat{\beta}_{OLS}$ and $\widehat{\beta}_{WLS}$.

nT	b						
	-0.5	-0.3	-0.1	0	0.1	0.3	0.5
450	1.4330	1.2521	1.0464	1.0000	1.0243	1.2314	1.6542
750	2.2503	1.5063	1.0438	1.0000	1.0368	1.4917	2.6015
1050	2.7276	1.9521	1.2325	1.0000	1.7976	2.1648	2.3052
1500	3.0328	2.5729	2.1721	1.0000	2.2849	2.9441	3.2411

Table 2. Ratio of mean-squared errors of $\widehat{\beta}_{OLS}$ and $\widehat{\widehat{\beta}}_n$.

nT	b						
	-0.5	-0.3	-0.1	0	0.1	0.3	0.5
450	2.5411	2.2252	2.1047	1.5594	2.0885	2.2441	2.8215
750	2.9576	3.6230	2.3868	1.7877	2.2245	2.4062	3.0274
1050	3.1926	2.6690	2.2065	2.0332	2.3264	2.8571	3.2024
1500	4.1032	3.7039	2.5880	2.1020	2.4210	3.8110	3.9507

Table 3. Ratio of mean-squared errors of $\widehat{\beta}_{WLS}$ and $\widehat{\widehat{\beta}}_n$.

nT	b						
	-0.5	-0.3	-0.1	0	0.1	0.3	0.5
450	1.7732	1.7771	2.0113	1.5594	2.0389	1.8223	1.7056
750	1.3143	2.4052	2.2866	1.7877	2.1455	1.6130	1.1637
1050	1.1704	1.3672	1.7902	2.0332	1.2941	1.3197	1.3892
1500	1.3529	1.4395	1.1914	2.1020	1.0595	1.2944	1.2189

Inspection of Tables: Each of the above simulations was replicated 200 times and the values of $\widehat{\beta}_{OLS}$, $\widehat{\beta}_{WLS}$ and $\widehat{\widehat{\beta}}_n$ were computed. Tables 1, 2 and 3 summarize the results for the ratio of mean squared errors of ($\widehat{\beta}_{OLS}$ and $\widehat{\beta}_{WLS}$), ($\widehat{\beta}_{OLS}$ and $\widehat{\widehat{\beta}}_n$) and ($\widehat{\beta}_{WLS}$ and $\widehat{\widehat{\beta}}_n$) respectively. It is seen in Monte Carlo studies that $\widehat{\beta}_{WLS}$ and $\widehat{\widehat{\beta}}_n$ are always more efficient than $\widehat{\beta}_{OLS}$. In addition to that, their efficiency versus $\widehat{\beta}_{OLS}$ increases as $|b|$ increase. Furthermore, in Table 1 when $b = 0$ the OLS and WLS estimators coincide. Also, Table 3 shows that the WLS estimator is dominated by the adaptive estimator.

4. Conclusions

The problem of estimating the coefficients of a multiple regression model in panel data with bilinear correlated errors is considered. The adaptive estimator and its asymptotic distribution are derived from the uniform local asymptotic normality of the model when the error terms are a bilinear process.

Furthermore, OLS and WLS estimators are derived and their limit distributions are studied. Finally, the simulation study shows that the adaptive estimators dominate both the weighted and unweighted (ordinary) least squares estimators.

Acknowledgements

The authors would like to thank the Editor and the Reviewers for their time, careful reading and comments which greatly improved the manuscript.

[1] Lillard L. A., Willis R. J. Dynamic aspects of learning mobility. Econometrica. **46** (5), 985–1012 (1978).

[2] Bhargava A., Franzini L., Narendranathan W. Serial correlation and the fixed effects model. The Review of Economic Studies. **49** (4), 533–549 (1982).

[3] Nicholls D. F., Pagan A. R., Terrell R. D. The estimation and use of models with moving average disturbance terms: A survey. International Economic Review. **16** (1), 113–134 (1975).

[4] Abonazel M. R. Different estimators for stochastic parameter panel data models with serially correlated errors. Journal of Statistics Applications & Probability. **7** (3), 423–434 (2018).

[5] Baltagi B., Li Q. Testing AR(1) against MA(1) disturbances in an error component model. Journal of Econometrics. **68** (1), 133–151 (1995).

[6] Allal J., El Melhaoui S. Tests de rangs adaptatifs pour les modèles de régression linéaire avec erreurs ARMA. Annales des Sciences Mathématiques du Québec. **30**, 29–54 (2006).

[7] Dutta H. Large sample tests for a regression model with autoregressive conditional heteroscedastic errors. Communications in Statistics – Theory and Methods. **28** (1), 105–117 (1999).

[8] Elmezouar Z. C., Kadi A. M., Gabr M. M. Linear regression with bilinear time series errors. PanAmerican Mathematical Journal. **22** (1), 1–13 (2012).

[9] Hallin M., Taniguchi M., Serroukh A., Choy K. Local asymptotic normality for regression models with long-memory disturbance. The Annals of Statistics. **27** (6), 2054–2080 (1999).

[10] Hwang S. Y., Basawa I. V. Asymptotic optimal inference for a class of nonlinear time series models. Stochastic Processes and their Applications. **46** (1), 91–113 (1993).

[11] Ling S., Peng L., Zhu F. Inference for a special bilinear time series model. Journal of Time Series Analysis. **36** (1), 61–66 (2015).

[12] Jiang J. Linear and Generalized Linear Mixed Models and Their Applications. Springer, New York (2007).

[13] Lmakri A., Akharif A., Mellouk A. Optimal detection of bilinear dependence in short panels of regression data. Revista Colombiana de Estadística. **43** (2), 143–171 (2020).

[14] Drost F. C., Klaassen C. A. J. Efficient estimation in semiparametric GARCH models. Journal of Econometrics. **81** (1), 193–221 (1997).

[15] Ling S., McAleer M. Adaptive estimation in nonstationry ARMA models with GARCH noises. The Annals of Statistics. **31** (2), 642–674 (2003).

[16] Le Cam L., Yang G. L. Asymptotics in Statistics. Springer, US (1990).

[17] Hájek J. A. Characterization of limiting distributions of regular estimates. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete. **14**, 323–330 (1970).

[18] Rao C. R. Linear Statistical Inference and Its Applications. Wiley, New York (1965).

[19] Schick A. On efficient estimation in regression models. The Annals of Statistics. **21** (3), 1486–1521 (1993).

[20] Koul H. L., Schick A. Efficient estimation in nonlinear autoregressive time series models. Bernoulli. **3**, 247–277 (1997).

Оцінка в короткопанельних моделях даних з білінійними помилками

Лмакрі А.[1], Ахаріф А.[2], Меллук А.[3]

[1] *Національна вища школа мистецтв і ремесел (ENSAM), Університет Хасана II, Касабланка, Марокко*
[2] *Лабораторія математики та прикладної математики FSTT, Університет Абдельмалек Ессааді Тетуан, Марокко*
[3] *Регіональний центр освіти та професійної підготовки, Танжер, Марокко*

У роботі розробляється асимптотична теорія для оцінювання в моделях коротких панельних даних з білінійною помилкою. Запропоновано порівняльне дослідження шляхом моделювання між декількома оцінками (адаптивними, звичайними та зваженим методом найменших квадратів) для коефіцієнтів моделей панельних даних, коли помилки є білінійними послідовно корельованими. Як наслідок властивості рівномірної локальної асимптотичної нормальності отримано адаптивні оцінки параметрів. Нарешті, проілюстровано продуктивність запропонованих оцінювачів за допомогою моделювання методом Монте–Карло. Показано, що адаптивні оцінки ефективніші, ніж зважені та звичайні оцінки методом найменших квадратів.

Ключові слова: *адаптивна оцінка; білінійні моделі; панельні регресійні моделі; зважені найменші квадрати; звичайні найменші квадрати.*