A Alternative Fixed Effects Estimator with AR(1) Errors

When the autoregressive coefficient, ρ, equals one, removal of the fixed effects via first differencing is the correct choice. If $\rho \in (-1, 1)$, however, it is possible that the fixed effects estimator could be a more appropriate choice.[1] Indeed, if the covariates are not strictly exogenous due to, say, measurement error, the first differences estimator will not be consistent. Even if the covariates are strictly exogenous, there is no practical method for determining *a priori* which method is best.

In light of these considerations, we also examined the results based on a fixed effects estimator.[2] The model is identical to before, with the exception that we now impose $\rho \in (-1, 1)$, thereby ruling out the unit root case. To be precise, we have

$$\begin{gathered} y_{it} = \mu_i + \gamma t + \beta' \mathbf{x}_{it} + e_{it} \\ e_{it} = \rho e_{i,t-1} + \varepsilon_{it}, \qquad \rho \in (-1, 1) \\ \varepsilon_{it} \sim N\left(0, \sigma^2\right) \\ t = 0, 1, 2, \ldots, T; \qquad i = 1, 2, 3, \ldots, N. \end{gathered}$$

As before, we put $\delta \equiv (\gamma, \beta')$. The joint density of system i's trip sequence can be then be written as

$$\begin{aligned} f_{\mathbf{y}_i}\left(\mathbf{y}_i\right) &= (2\pi)^{-(T+1)/2} \left|\mathbf{\Sigma}\right|^{-1/2} \\ &\quad \cdot \exp\left(-\frac{1}{2}\left(\mathbf{y}_i - \boldsymbol{\iota}_{T+1}\mu_i - \mathbf{Z}_i\boldsymbol{\delta}\right)' \mathbf{\Sigma}^{-1}\left(\mathbf{y}_i - \boldsymbol{\iota}_{T+1}\mu_i - \mathbf{Z}_i\boldsymbol{\delta}\right)\right) \quad (1) \end{aligned}$$

where

$$\begin{gathered} \mathbf{y}_i \equiv \left(y_{i0},\ y_{i1},\ \ldots,\ y_{iT}\right)', \\ \boldsymbol{\iota}_{T+1} \text{ is a } (T+1)\text{-vector of ones}, \\ \mathbf{Z}_i = \begin{bmatrix} 1 & \mathbf{x}'_{i0} \\ 2 & \mathbf{x}'_{i1} \\ \vdots & \\ T+1 & \mathbf{x}'_{iT} \end{bmatrix}, \end{gathered}$$

[1] Assuming, of course, the number of time periods exceeds two. When there are just two time periods, the fixed effects and first difference approaches are identical.

[2] Detailed results are available from us on request.

and

$$\boldsymbol{\Sigma} = \frac{\sigma^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^T \\ \rho & 1 & \rho & \cdots & \rho^{T-1} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^T & \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{bmatrix}.$$

Since systems are independent, the log likelihood for the entire panel is merely

$$\begin{aligned} L(\boldsymbol{\varphi}) &= \sum_{i=1}^{N} \ln f_{\mathbf{y}_i}(\mathbf{y}_i) \\ &= \sum_{i=1}^{N} \Bigg(-\frac{T+1}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}| \\ &\quad -\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\iota}_{T+1}\mu_i - \mathbf{Z}_i\boldsymbol{\delta})' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\iota}_{T+1}\mu_i - \mathbf{Z}_i\boldsymbol{\delta}) \Bigg), \end{aligned}$$

where $\boldsymbol{\varphi} \equiv (\mu_1,\ \mu_2,\ \ldots,\ \mu_N,\ \boldsymbol{\delta}',\ \rho,\ \sigma^2)'$.

As before, we let $\mathbf{A}_i$ be a $T_i^* \times (T+1)$ matrix such that the non-missing trips for system i are given by

$$\mathbf{y}_i^* \equiv \mathbf{A}_i \mathbf{y}_i, \qquad i = 1, 2, 3, \ldots, N.$$

The log likelihood for our observed (i.e., non-missing) data is then

$$\begin{aligned} L^*(\boldsymbol{\varphi}) &= \sum_{i=1}^{N} \Bigg(-\frac{T_i^*}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i^*| \\ &\quad -\frac{1}{2} \left(\mathbf{y}_i^* - \boldsymbol{\iota}_{T_i^*}\mu_i - \mathbf{Z}_i^*\boldsymbol{\delta}\right)' (\boldsymbol{\Sigma}_i^*)^{-1} \left(\mathbf{y}_i^* - \boldsymbol{\iota}_{T_i^*}\mu_i - \mathbf{Z}_i^*\boldsymbol{\delta}\right) \Bigg), \end{aligned} \tag{2}$$

where $\mathbf{Z}_i^* \equiv \mathbf{A}_i \mathbf{Z}_i$ and $\boldsymbol{\Sigma}_i^* \equiv \mathbf{A}_i \boldsymbol{\Sigma} \mathbf{A}_i'$.

Direct numerical maximization of (2) is problematic due to the dimensionality of $\boldsymbol{\varphi}$. Since the maximum likelihood estimators (MLEs) of the fixed effects conditional on $\boldsymbol{\theta} \equiv (\boldsymbol{\delta}',\ \rho,\ \sigma^2)'$ are given by

$$\widehat{\mu}_i \equiv \widehat{\mu}_i(\boldsymbol{\theta}) = \left(\boldsymbol{\iota}_{T_i^*}' (\boldsymbol{\Sigma}_i^*)^{-1} \boldsymbol{\iota}_{T_i^*}\right)^{-1} \boldsymbol{\iota}_{T_i^*}' (\boldsymbol{\Sigma}_i^*)^{-1} (\mathbf{y}_i^* - \mathbf{Z}_i^*\boldsymbol{\delta}),$$

Time	Log of Total Population	Log of Population Aged 65 and Over	ρ
0.071^{**}	0.574	0.508	0.301^{**}
(0.009)	(0.446)	(0.462)	(0.035)

Table 1: Maximum Likelihood Estimates for the FE Model (Standard errors in parentheses. Double starred items are significant at the 0.01 level.

this problem can be simplified by maximizing the concentrated log likelihood function

$$\begin{aligned} L_c^*(\boldsymbol{\theta}) &= \sum_{i=1}^{N} \left(-\frac{T_i^*}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i^*| \right. \\ & \left. -\frac{1}{2} \left(\mathbf{y}_i^* - \boldsymbol{\iota}_{T_i^*} \widehat{\mu}_i - \mathbf{Z}_i^* \boldsymbol{\delta} \right)' (\boldsymbol{\Sigma}_i^*)^{-1} \left(\mathbf{y}_i^* - \boldsymbol{\iota}_{T_i^*} \widehat{\mu}_i - \mathbf{Z}_i^* \boldsymbol{\delta} \right) \right) \end{aligned} \quad (3)$$

with respect to $\boldsymbol{\theta}$ and then using the invariance of the MLE to obtain the MLEs of the fixed effects.

Table 1 presents the maximum likelihood estimates when all covariates are included in the model. As with the first differenced version, the time trend and the autoregressive parameter are statistically significant while the population variables are insignificant. While the estimated trend coefficients are similar in magnitude (0.071 vs. 0.079), the estimated autoregressive coefficients differ substantially (0.301 vs. 0.948).

Obtaining the standard error for the MLE of ρ in the fixed effects version is computationally intensive since it requires either computing the numerical hessian of (2) with respect to $\boldsymbol{\varphi}$ or the outer product of the gradient of the log of (1) with respect to $\boldsymbol{\varphi}$ for each system. However, using the invariance principle, the MLEs and standard errors of $\boldsymbol{\delta}$ can be computed relatively inexpensively via GLS by using just the MLE of ρ (which can be found from the concentrated log likelihood function). Table 2 reports results for the remaining combinations of covariates.

Time	Log of Total Population	Log of Population Aged 65 and Over
0.084** (0.004)		
	4.246** (0.350)	
		3.255** (0.077)
0.077** (0.005)	0.888** (0.324)	
0.064** (0.009)		0.859* (0.342)
	-0.778 (0.411)	3.398** (0.226)

Table 2: Maximum Likelihood Coefficient Estimates for Other Combinations of Covariates in the FE Model (Standard errors in parentheses. Double starred items are significant at the 0.01 level.