A note on the estimation of AR(1) fixed-effects regressions in unbalanced panels.

Alexandre Cazenave-Lacroutz[*,a,b], Vieu Lin[*]

October 29, 2019

Version 1.0.6 - comments are welcome !

Abstract

This note discusses the estimation of the parameters of fixed-effects models with AR(1) perturbations. First, it shows the Baltagi et Wu (1999) transformation is not perfectly adapted in this case, and suggests a modification of this transformation that provides a consistent estimation of β (that may be more efficient than estimating β by ignoring the autocorrelation of the perturbations). Second, it highlights that the variance of the perturbations is very badly estimated by the current estimation procedure, and proposes a method to consistently estimate it. Monte Carlo simulations illustrate the properties of these new estimation methods.

[*]Institut National de la Statistique et des Études Économiques, 88 Avenue Verdier, Montrouge. This document does not reflect the position of Insee, Université Paris Dauphine or Crest, but only its authors' views. This research was conducted while implementing new wage equations in the Insee microsimulation model Destinie 2, see Cazenave-Lacroutz *et al.* (2019). Alexandre thanks Prof.Dr. Winter for having hosted him at LMU University of Munich when this research started.

[a]Université Paris Dauphine, Place du Maréchal de Lattre de Tassigny, 75016 Paris

[b]Crest, 5 Avenue Henry Le Chatelier, 91120 Palaiseau

1 Introduction

This note considers the estimation of linear unobserved effects panel data models[1] with AR(1) disturbances, that refer to processes of the form:

$$y_{it} = x'_{it}\beta + \nu_i + u_{it} \tag{1}$$

$$u_{it} = \rho u_{it-1} + \varepsilon_{it} \tag{2}$$

where $|\rho| < 1$ and the ε_{it}'s are i.i.d. disturbances with mean 0 and variance σ_ε^2.

Such estimations are common in the wage equation literature (with even higher order of correlation), and are not infrequent in the general economics literature. They are performed in Stata with the command *xtregar*, which has been used in influential and recent economics articles such as Dafny (2010), Hau *et al.* (2013) or Prieto et Lago-Peñas (2012).

In Cazenave-Lacroutz et Lin (2019), we proposed new estimators for the auto-correlation coefficient. In the particular case of random-effects setting (that is: ν_i is exogenous to the covariates) and a positive auto-correlation parameter[2], it enables to produce consistent estimators for all parameters of the model, assuming there are enough individuals with at least two consecutive observations and at least three observations overall.

We now turn to the more general case when no hypothesis is made regarding the exogeneity of the unobserved individual effect. First we remind the properties of the current estimation method implemented in the Stata command **xtregar, fe**. Additionnaly, we explain why it works very well to estimate the variance of the perturbation in the balanced case, but fails to do so in the unbalanced case. We also suggest a novel estimation method for the coefficients of the independent variables, as the current estimation method may perform poorly in some circumstances. Second, a consistent estimator of the variance of the perturbations is derived. Third, we illustrate the properties of these estimators though Monte Carlo simulations that

[1] That is in both fixed and random effects models.

[2] In case of a negative auto-correlation parameter, asymptotically consistent estimator is provided, where the convergence is achieved when the minimum number of period per individual goes to infinity.

compare it to the (more specific) random-effects estimators and to the (more general) fixed-effects estimators. We eventually conclude.

2 The current estimation method

In all the paper, we consider that a consistent estimator of the autocorrelation parameter ρ is known (Cazenave-Lacroutz et Lin, 2019). To alleviate the notations, we even consider that the true value of ρ is known.

When the fixed-effects are assumed to be exogeneous to the covariates (the random-effects model), Baltagi et Wu (1999) propose a transformation of the data that enables to get consistent estimates of the β parameter, of the variance of the fixed-effects, and of the variance of the perturbations.

In the general case (when there is no hypothesis regarding the exogeneity of the fixed-effects), the current estimation method (for instance implemented by the Stata command xtregar) first applies to the fixed-effects model the Baltagi-and-Wu transformation (Baltagi et Wu, 1999). We detail it below.

2.1 The current estimation method

Baltagi et Wu (1999) derive a transformation of the data that removes the AR(1) component. Following this transformation, one gets:

$$\begin{aligned} y^*_{i,t_{i,j}} &= (1-\rho^2)^{1/2} y_{i,t_{i,j}} \text{ if } j=1 \\ &= (1-\rho^2)^{1/2} (y_{i,t_{i,j}} \frac{1}{(1-\rho^{2(t_{i,j}-t_{i,j-1})})^{1/2}} - y_{i,t_{i,j-1}} \frac{\rho^{t_{i,j}-t_{i,j-1}}}{(1-\rho^{2(t_{i,j}-t_{i,j-1})})^{1/2}}) \text{ if } j>1 \end{aligned} \tag{3}$$

Let us consider equation (1):

$$y_{it} = x'_{it}\beta + \nu_i + u_{it}$$

By applying the Baltagi-and-Wu transformation, one gets:

$$y^*_{it} = x^{*\prime}_{it}\beta + \nu^*_{i,t} + u^*_{it} \tag{4}$$

Quite importantly, note that the fixed effects ν_i have become $\nu^*_{i,t}$. In the general case, the Baltagi-transformed fixed effects are no longer fixed over time. The balanced case is almost an exception to that regard, as in this very particular case:

$$\begin{aligned} v^*_{it} &= (1-\rho^2)^{1/2}\nu_i \text{ if } t = 1 \\ &= (1-\rho)\nu_i \text{ if } t > 1 \end{aligned} \tag{5}$$

It is easy to show that the transformed error terms u^*_{it} are no longer correlated and are homoskedastic.

The current estimation method (see Stata - xtregar / Methods and Formula) then substracts to this equation the mean and the grand mean. For a given variable x, we define the following quantities, where n_i is equal to the number of occurrences of unit i :

$$\overline{x^*} = \frac{\sum_{j=2}^{n_i} x^*_{i,t(i,j)}}{n_i - 1}$$

$$\overline{\overline{x^*}} = \frac{\sum_{i=1}^{N}\sum_{j=2}^{n_i} x^*_{i,t(i,j)}}{\sum_{i=1}^{N}(n_i - 1)}$$

$$x^{**}_{it} = x^*_{it} - \overline{x^*} + \overline{\overline{x^*}} \tag{6}$$

The transformed equation is thus:

$$y^{**}_{it} = x^{**'}_{it}\beta + \nu^{**}_{it} + u^{**}_{it} \tag{7}$$

An OLS regression of y^{**} on x^{**} and a constant is then performed. This method would provide a consistent estimate of β if ν^*_{it} did not depend on t.

2.2 The estimation of the β

Even if this method seems to provide consistent estimators of β in many of our Monte Carlo simulations, we are able to show its consistency in the unbalanced case only under specific assumptions (see Annex A). In Section 5, we also present a Monte Carlo simulation where β does not seem to be consistently estimated[3].

[3]In the balanced case, it would provide a consistent estimates if the first observations were not used, as all the other $\nu^*_{i,t}$ are fixed over time.

Happily, consistent estimators of β already exist. Moreover, we propose in Section 3 a novel (consistent) estimator of β that takes into account the autocorrelation in the perturbations.

2.3 An imprecise estimation of the fixed effects ν_i

Once a consistent estimator of β has been computed, the parameters ν_i can be estimated by:

$$\hat{\nu}_i = \overline{y}_i - \overline{x_i}'\hat{\beta} \tag{8}$$

These estimates might be made centered around zero. Even without centering it, the empirical variance of these estimates provides an estimate of the variance of the fixed-effects. As such variance is based on the imprecise estimates of the individual fixed-effects, it is quite imprecise. It converges however towards the true variance when the minimal number of observation per individual tends to infinity.

2.4 The variance of the perturbations

We focus here on the current estimation of the variance of the perturbations σ_ε^2, implemented in Stata command xtregar. One estimation method consists in considering the empirical variance of $u_{i,t}^*$. We explain below why it seems to yield a consistent estimator of σ_ε^2. But in the unbalanced case, there is no reason why it would yield a consistent estimator of σ_ε^2. Monte Carlo simulations in Section 5 clearly shows that the estimates obtained from the current command **xtregar** can be far away from the true value of σ_ε^2.

The balanced case:

By applying the Baltagi-and-Wu transformation to the balanced case to the perturbation $u_{i,t}$, it comes:

$$\begin{aligned} u^*_{i,t} &= (1-\rho^2)^{1/2} u_{i,1} \text{ if } t = 1 \\ &= u_{i,t} - \rho u_{i,t-1} \text{ if } t > 1 \end{aligned}$$

We remind equation (2):

$$u_{i,t} = \rho u_{i,t-1} + \varepsilon_{it}$$

Hence, $u^*_{i,t} = \varepsilon_{it}$ if $t > 1$. It comes:

$$\begin{aligned} \mathrm{Var}(u^{**}_{it}) &= (1 + \frac{N-1}{N(T-1)}(\frac{2}{T-1} + \frac{1}{N} - 3))\sigma^2_\varepsilon \\ &\neq \sigma^2_\varepsilon \end{aligned}$$

3 Consistent estimators of β

A first method to get a consistent estimator of β consists in estimating the model without taking into account the autocorrelation in the perturbations. That is: one estimates equation (1) without taking into account equation (2)[4]. One needs to adjust the computation of the standard errors of the estimated coefficients to account for the presence of serial correlation within each individual unit[5].

As a second (natural) method of estimation of β, we propose a modification of the Baltagi-and-Wu transformation that yields another consistent estimator of the β parameter. In some circumstances, it delivers a more efficient estimator of the β than with the first method of estimation above.

We observe that:

$$y^*_{it} = x^{*'}_{it}\beta + (1-\rho^2)^{1/2} 1_{i,t}\nu_i + u^*_{it}$$

[4]There is at least two possibilities that can be considered: the so-called "fixed-effects" estimator, or the "first-difference" estimator. As explained by Wooldridge (2010) (in section 10.7.1), the first one is more efficient than the second one when errors terms are i.i.d. (that is $\rho = 0$) while the inverse is true when errors terms follow a random walk (that is $\rho = 1$). Which one is more efficient in the other cases is not trivial. To ease the exposition, we consider in the following only the fixed-effects estimator.

[5]In practice, in Stata, one should use: **xtreg, fe vce(cluster id)**. Alternatively, once β have been consistently estimated, section 4 provides a consistent estimator of σ^2_u which can in turn be used to estimate the variance associated with the estimation of $\hat{\beta}$.

Where:

$$\begin{aligned} 1_{i,t_{i,j}} &= 1 \text{ if } j = 1 \\ &= \frac{1 - \rho^{t_{i,j} - t_{i,j-1}}}{(1 - \rho^{2(t_{i,j} - t_{i,j-1})})^{1/2}} \text{ if } j > 1 \end{aligned} \tag{9}$$

Thus, we propose the following transformation that makes the transformed fixed-effects to remain fixed over time, and keeps the transformed perturbations uncorrelated:

$$x_{it}^{*2} = \frac{1}{1_{i,t}} x_{it}^{*} \tag{10}$$

Hence, our modification of the Baltagi-and-Wu transformation writes:

$$\begin{aligned} x_{i,t_{i,j}}^{*2} &= (1 - \rho^2)^{1/2} x_{i,t_{i,j}} \text{ if } j = 1 \\ &= (1 - \rho^2)^{1/2} \left(x_{i,t_{i,j}} \frac{1}{1 - \rho^{t_{i,j} - t_{i,j-1}}} - x_{i,t_{i,j-1}} \frac{\rho^{t_{i,j} - t_{i,j-1}}}{1 - \rho^{t_{i,j} - t_{i,j-1}}} \right) \text{ if } j > 1 \end{aligned} \tag{11}$$

From that point on, the principle is the following: by demeaning as in the current method, we get rid of all the terms with ν_i:

$$y_{i,t}^{*2} - \overline{y_i^{*2}} = (x_{i,t}^{*2\prime} - \overline{x_i^{*2}}^{\prime})\beta + (u_{i,t}^{*2} - \overline{u_i^{*2}})$$

An OLS estimation of the above regression yields a consistent estimate of β.

In practice, the demeaning is performed using Stata **xtreg, fe vce(robust)** command, as it handles the remaining heteroskedasticity and serial correlation in the transformed errors u_{it}^{*2}. The above Stata command indeed takes into account the note of Stock et Watson (2008) that tackles both issues[6].

[6] Note that the variance-covariance matrix of the error terms $u_{i,t}^{*2}$ is a diagonal matrix whose diagonal coefficients are $\frac{\sigma_\varepsilon^2}{1_{i,t}^2}$. As it is explicitly known, a more efficient estimator of β that takes into account this explicit form may exist. This is let to further work.

A comparison of the two methods

In the general case, none of the method presented above is necessarily more efficient than the other. Which one is more efficient depends on the pattern of the missing observations, the autocorrelation coefficient ρ and the covariates x_{it} considered.

The first method estimator $\hat{\beta}_1$ is simply the fixed-effects estimator obtained from the following equation:

$$y_{it} = x'_{it}\beta + \nu_i + u_{it}$$

The second method estimator $\hat{\beta}_2$ is the fixed-effects estimator obtained from the following equation:

$$y_{it}^{*2} = x_{it}^{*2'}\beta + \nu_i^{*2} + u_{it}^{*2}$$

Both equations include an individual effect. However, in the first equation, the idiosyncratic errors are are homoskedastic and serially correlated, whereas they are heteroskedastic but not serially correlated in the second equation.

Let us note e_1 and e_2 the error term after applying the within transformation to the first and second equation. We also denote with a tilde a variable whose individual mean has been substracted to. We have:

$$V(\hat{\beta}_1|X) = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'V(e_1|X)\tilde{X}(\tilde{X}'\tilde{X})^{-1}$$

$$V(\hat{\beta}_2|X) = (\widetilde{X^{*2'}}\widetilde{X^{*2}})^{-1}\widetilde{X^{*2'}}V(e_2|X)\widetilde{X^{*2}}(\widetilde{X^{*2'}}\widetilde{X^{*2}})^{-1}$$

These variances are not easy to compare. First and foremost, $V(\hat{\beta}_2|X)$ depends non-linearly on $X^{*2'}$, that is itself a transformation involving ρ. Second, even in the simple balanced case with a long time dimension, the central term are not easy to compare. $V(e_1)$ is similar to a matrix $M_{(i,p),(i,s)}$ whose cross-individual terms are null, but whose terms over individuals are: $\rho^{p-s}\frac{\sigma_\epsilon^2}{1-\rho^2}$. $V(e_2)$ is close to a diagonal matrix whose diagonal terms are σ_ϵ^2.

We provide Monte Carlo simulations where the first method of estimation is more efficient than the second (e.g. Table 1) and other where the reverse holds. As a general advice, we suggest to the interested researchers to perform the two analyses[7]

[7]To be complete, β should also be estimated with the first-differences estimator.

and to choose the one whose reported standard errors of β are the lowest.

4 A new estimator of the variance of the perturbations

A naive estimator of σ_ε^2 can be obtained by explicitly computing $\hat{u}_{i,t} := y_{i,t} - x'_{i,t}\hat{\beta} - \hat{\nu}_i$; considering its empirical variance, and multiplying it by $1-\rho^2$ to obtain an estimator σ_ε^2. It however yields an imprecise estimator of σ_ε, as it relies on the imprecise estimation of the ν_i. We therefore propose an estimator that is not based on the estimation of the ν_i.

To do so, we define:

$$\widetilde{y}_{i,t} = y_{i,t} - x'_{i,t}\beta \tag{12}$$

As highlighted above, under the missing-at-random hypothesis, the current estimation method enables to get a consistent estimate of β, and thus to compute a consistent estimate of the above quantity.
We observe that:

$$\widetilde{y}_{i,t} = y_{i,t} - x'_{i,t}\beta = \nu_i + u_{i,t} \tag{13}$$

We define :

$$\widetilde{\widetilde{y}}_{i,t(i,j)} = \widetilde{y}_{i,t(i,j)} - \widetilde{y}_{i,t(i,j-1)} = u_{i,t(i,j)} - u_{i,t(i,j-1)} \tag{14}$$

By successively applying equation (2), it comes:

$$\widetilde{\widetilde{y}}_{i,t(i,j)} = (\rho^{t(i,j)-t(i,j-1)} - 1)u_{i,t(i,j-1)} + \sum_{k=0}^{t(i,j)-t(i,j-1)-1} \rho^k \varepsilon_{i,t(i,j)-k} \tag{15}$$

Hence, since all the above terms in the sum are uncorrelated and centered:

$$\mathrm{E}(\widetilde{\widetilde{y}}\,^2_{i,t(i,j)}) = \mathrm{Var}(\widetilde{\widetilde{y}}_{i,t(i,j)}) = (1-\rho^{t(i,j)-t(i,j-1)})^2\sigma_u^2 + \sum_{k=0}^{t(i,j)-t(i,j-1)-1} \rho^{2k}\sigma_\varepsilon^2$$

Thus:

$$\sigma_\varepsilon^2 = \frac{E(\widetilde{\widetilde{y}}\,^2_{i,t(i,j)})}{\frac{(1-\rho^{t(i,j)-t(i,j-1)})^2}{(1-\rho^2)} + \frac{(1-\rho^{2(t(i,j)-t(i,j-1))})}{(1-\rho^2)}} \tag{16}$$

We have obtained:

$$\sigma_\varepsilon^2 = \mathrm{E}(w_{it_{ij}}) \tag{17}$$

where:

$$w_{it_{ij}} = \frac{\widetilde{\widetilde{y}}\,^2_{i,t(i,j)}}{\frac{(1-\rho^{t(i,j)-t(i,j-1)})^2}{(1-\rho^2)} + \frac{(1-\rho^{2(t(i,j)-t(i,j-1))})}{(1-\rho^2)}} \tag{18}$$

Under very general hypotheses, one gets a natural consistent estimate of σ_ε^2 (see Annex B for a proof) :

$$\widehat{\sigma_\varepsilon^2} = \frac{1}{N}\sum_{i=1}^{N} \overline{w}_i \tag{19}$$

5 Monte Carlo simulations

To illustrate the previous theoretical sections, we perform Monte Carlo simulations with the following basis parameters : $N = 500$, $T = 10$, $\rho = 0.6$, $\sigma_\varepsilon = 0.3$, $\sigma_\nu = 0.35$. Those parameters were already used by Cazenave-Lacroutz et Lin (2019), enabling comparability [8].

5.1 In the random effects design

First, we consider the case where the individual effects ν_i are exogeneous from the covariables x[9]. This is a very particular case, where a random-effects estimation method can also be applied. This allows to compare the estimation advantages of making the assumption of a random-effects model (when suitable) rather than the more general fixed-effects model.

With both random-effects and fixed-effects models, we note that the estimation of β is (slightly) more efficient with **xtreg** than the one we proposed. In the general fixed-effects case, we also note that in this particular case, the estimation of β by xtregar does not seem biased, although it is far less efficient than our method of

[8]The values of ρ, σ_ε and σ_ν were chosen by Cazenave-Lacroutz et Lin (2019), as they were typical of what they encountered in an applied study, see Cazenave-Lacroutz *et al.* (2019). The values $T = 10$ and $N = 500$ ensure that the panel considered is short in the time dimension but with an number of individual units close to infinity.

[9]Typically, we consider covariables as a random draw that has an additive impact on the dependent variable, and a constant.

estimation or the one of **xtreg**.

With our estimation method for σ_ϵ, both the fixed-effects method and the random-effects method provide consistent estimates of the standard deviation of the perturbations (see Table 1). However, due to the limited number of periods observed per individual, the estimation of the standard deviation σ_ν derived from the fixed effects method is biased, but unbiased with the random-effect method[10].

We also note that in this particular case, the coefficient of the covariable is consistently estimated by both estimation methods.

In Table 2, the same simulations are performed with a longer time period ($T = 100$ rather than $T = 10$). Accordingly with our previous analysis, the fixed-effects method yields estimates of the standard deviation σ_ν that are no longer significantly different from its true value. Unlike what was observed in Table 1, with both random-effects and fixed-effects models, we also note that the estimation of β is (slightly) more efficient with our method of estimation than with **xtreg**.

[10]Indeed, in the random-effects method, it can be estimated without relying on the imprecise estimation of the various ν_i.

Table 1. Monte Carlo simulations on an unbalanced panel, $T = 10$

	Fixed-effects method				Random-effects method			
	ρ	σ_ε	σ_ν	*covar*	ρ	σ_ε	σ_ν	*covar*
true values	0.6	0.3	0. 35	3	0.6	0.3	0. 35	3
with true ρ and xtregar	.6 (0)	.621*** (4.9e-03)	.447*** (.013)	3 (.01)	.6 (0)	.3 (4.2e-03)	.356 (.016)	3 (6.4e-03)
with true ρ and corrected xtregar	.6 (0)	.3 (4.6e-03)	.447*** (.013)	3 (6.7e-03)	.6 (0)	.3 (4.2e-03)	.356 (.016)	3 (6.4e-03)
with ρ_{BFN} and corrected xtregar	.581 (.04)	.299 (3.2e-03)	.447*** (.013)	3 (6.5e-03)	.581 (.04)	.3 (3.9e-03)	.359 (.023)	3 (6.3e-03)
with true ρ and xtreg	0 (0)	.323*** (5.9e-03)	.432*** (.013)	3 (5.2e-03)	0 (0)	.323*** (5.9e-03)	.403*** (.014)	3 (5.1e-03)

Legend: The average estimators should not be significantly different from the true values. It is the case only for those in bold. Significance levels for the differences with the true values are otherwise pinpointed by stars: * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$)
Note 1: Approximately half of a panel of 500 individuals observed each over T = 10 periods has been randomly deleted, before the Monte Carlo process has been implemented with 50 replications.
The estimates of σ_ε and of σ_ν in the two last lines are obtained by estimating first ρ_{BFN} (or ρ_{BFN2U}), and then by imposing it as the estimate of ρ in *xtregar*.

Table 2. Monte Carlo simulations on an unbalanced panel, $T = 100$

	Fixed-effects method				Random-effects method			
	ρ	σ_ε	σ_ν	*covar*	ρ	σ_ε	σ_ν	*covar*
true values	0.6	0.3	0. 35	3	0.6	0.3	0. 35	3
with true ρ and xtregar	.6 (0)	.678*** (2.8e-03)	.36 (.01)	3 (4.9e-03)	.6 (0)	.3 (1.1e-03)	.35 (9.9e-03)	3 (2.4e-03)
with true ρ and corrected xtregar	.6 (0)	.3 (1.3e-03)	.36 (.01)	3 (2.7e-03)	.6 (0)	.3 (1.1e-03)	.35 (9.9e-03)	3 (2.4e-03)
with ρ_{BFN} and corrected xtregar	.601 (4.6e-03)	.3 (1.2e-03)	.36 (.01)	3 (2.7e-03)	.601 (4.6e-03)	.3 (1.2e-03)	.35 (9.9e-03)	3 (2.4e-03)
with xtreg	0 (0)	.37*** (1.8e-03)	.36 (1.0e-02)	3 (3.0e-03)	0 (0)	.37*** (1.8e-03)	.356 (.01)	3 (3.0e-03)

Legend: The average estimators should not be significantly different from the true values. It is the case only for those in bold. Significance levels for the differences with the true values are otherwise pinpointed by stars: * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$)

Note 1: Approximately half of a panel of 500 individuals observed each over T = 100 periods has been randomly deleted, before the Monte Carlo process has been implemented with 50 replications.

The estimates of σ_ε and of σ_ν in the two last lines are obtained by estimating first ρ_{BFN} (or ρ_{BFN2U}), and then by imposing it as the estimate of ρ in *xtregar*.

5.2 In the general case

We introduce two changes in regard with the simulations presented in Table 1.
First, the individual effects are no longer exogeneous to the covariates[11]. Hence, we no longer present the results from the random-effects method, as it is based on this assumption of orthogonality.
Second, while the data were missing at random, we deviate from this missing pattern. In the "Non missing at random" setting, missingness is based on the value taken by the covariates. As shown in Table 3, this yields an estimate for the coefficient of the covariable that is significantly different from its true value, which was not the case under the Missing-at-random setting. Up to now, we do not know to which extent the absence of the Missing-at-random hypothesis is central in explaining this. Our main purpose is to show that the parameter β is not always consistently estimated under the current procedure. As expected, our correction procedure yields a consistent estimate in both studied cases.

[11] More precisely, the covariate is the sum of a random term and the fixed-effect.

Table 3. Monte Carlo simulations on an unbalanced panel, $T = 10$, general case

	Missing-at-random				Non missing-at-random			
	ρ	σ_ε	σ_ν	covar	ρ	σ_ε	σ_ν	covar
true values	0.6	0.3	0. 35	3	0.6	0.3	0. 35	3
with true ρ and xtregar	.6 (0)	.621*** (4.9e-03)	.442*** (.019)	3 (.012)	.6 (0)	.458*** (.017)	.377 (.021)	2.95*** (.013)
with true ρ and corrected xtregar	.6 (0)	.3 (4.6e-03)	.447*** (.016)	3 (6.7e-03)	.6 (0)	.302 (7.5e-03)	.355 (.018)	3 (8.0e-03)
with ρ_{BFN} and corrected xtregar	.581 (.04)	.299 (3.2e-03)	.447*** (.016)	3 (6.5e-03)	.581 (.04)	.3 (8.3e-03)	.355 (.018)	3 (8.0e-03)
with xtreg	0 (0)	.323*** (5.9e-03)	.434*** (.015)	3 (5.2e-03)	0 (0)	.326*** (7.0e-03)	.373 (.021)	2.99 (.013)

Legend: The average estimators should not be significantly different from the true values. It is the case only for those in bold. Significance levels for the differences with the true values are otherwise pinpointed by stars: * $(p < 0.10)$, ** $(p < 0.05)$, *** $(p < 0.01)$

Note 1: Approximately half of a panel of 500 individuals observed each over T = 10 periods has been randomly deleted, before the Monte Carlo process has been implemented with 50 replications.

The estimates of σ_ε and of σ_ν in the two last lines are obtained by estimating first ρ_{BFN} (or ρ_{BFN2U}), and then by imposing it as the estimate of ρ in *xtregar*.

Note 2: *covar* is an independent variable that is the sum of a random term and the fixed-effect.

6 Conclusion

Whereas Baltagi et Wu (1999) proposed a way of consistently estimating the parameters of a random-effects regression with AR(1) perturbations[12], to the best of our knowledge, no previous paper attempted to generalize their method to the case where no hypothesis is made regarding the exogeneity of the fixed effects. Current practice consisted in applying the Baltagi-and-Wu transformation before demeaning, but we show this is not adapted.

In the unbalanced case, we provide Monte Carlo simulations where the current procedure does not yield a consistent estimation of the β parameter. Furthermore, we propose a novel transformation that provides consistent estimates also in the unbalanced case (that may be more efficient than estimating β by ignoring the autocorrelation of the perturbations). We also notice that the variance of the perturbation was not consistently estimated. In case a consistent estimate of β is available, we propose an additional estimation procedure that enables to get a consistent estimator of σ_ϵ. Hence, all parameters of this type of model can now be consistently estimated[13].

References

Baltagi, B. H. et Wu, P. X. (1999). Unequally spaced panel data regressions with AR (1) disturbances. *Econometric Theory*, 15(6):814–823.

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

Cazenave-Lacroutz, A., Godet, F. et Lin, V. (2019). Modélisation des trajectoires de revenus d'activité pour le modèle destinie 2.

Cazenave-Lacroutz, A. et Lin, V. (2019). The estimation of the autocorrelation coefficient in panel data models with ar(1) disturbances.

Dafny, L. S. (2010). Are health insurance markets competitive? *American Economic Review*, 100(4):1399–1431.

[12]when a consistent estimator of the autocorrelation parameter ρ is available. Such a consistent estimator is proposed by Cazenave-Lacroutz et Lin (2019).

[13]The variance of the fixed-effects is consistently estimated as T tends to ∞.

Hau, H., Langfield, S. et Marques-Ibanez, D. (2013). Bank ratings: what determines their quality? *Economic Policy*, 28(74):289–333.

Prieto, D. C. et Lago-Peñas, S. (2012). Decomposing the determinants of health care expenditure: the case of spain. *The European Journal of Health Economics*, 13(1):19–27.

Stock, J. H. et Watson, M. W. (2008). Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica*, 76(1):155–174.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Annexes

A Consistency of $\hat{\beta}$

We focus on the consistency of $\hat{\beta}$, where $\hat{\beta}$ denotes the OLS estimator of β obtained from equation (7). It suffices to show that the orthogonality condition $\mathbb{E}(x^{**}_{it_{ij}}(\nu^{**}_{it_{ij}} + u^{**}_{it_{ij}})) = 0$ holds. First, the relation $\mathbb{E}(x^{**}_{it_{ij}} u^{**}_{it_{ij}}) = 0$ trivially follows from the strict exogeneity assumptions $\mathbb{E}(x_{it_{ij}} u_{kt_{kl}}) = 0$. According to Lemma 1, the second relation $\mathbb{E}(x^{**}_{it_{ij}} \nu^{**}_{it_{ij}}) = 0$ holds if one of two conditions holds.

Lemma 1: Condition $\mathbb{E}(x^{**}_{it_{ij}} \nu^{**}_{it_{ij}}) = 0$ holds for all $2 \leq j \leq n_i$
if:

$$\frac{1-\rho^{t_{ij}-t_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}} = (1 - \frac{n_i - 1}{\sum\limits_{i=1}^{N}(n_i - 1)})\frac{1}{n_i - 1}\sum_{l=2}^{n_i}\frac{1-\rho^{t_{il}-t_{il-1}}}{\sqrt{1-\rho^{2(t_{il}-t_{il-1})}}} \tag{20}$$

for all $2 \leq j \leq n_i$.

or if:

$$\mathbb{E}(x^{**}_{it_{ij}} \nu_i) = 0 \tag{21}$$

First, we compute

$$\begin{aligned}
\mathbb{E}(x^{**}_{it_{ij}} \nu^{*}_{it_{ij}}) &= \mathbb{E}(x^{**}_{it_{ij}} \sqrt{1-\rho^2}\frac{1-\rho^{t_{ij}-t_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}}\nu_i) \\
&= \sqrt{1-\rho^2}\frac{1-\rho^{t_{ij}-t_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}}\mathbb{E}(x^{**}_{it_{ij}} \nu_i),
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}(x^{**}_{it_{ij}}\overline{\nu^*_i}) &= \mathbb{E}(\frac{1}{n_i-1}\sum_{l=2}^{n_i} x^{**}_{it_{ij}}\nu^*_{it_{il}}) \\
&= \mathbb{E}(\frac{1}{n_i-1}\sum_{l=2}^{n_i}\sqrt{1-\rho^2}\frac{1-\rho^{t_{il}-t_{il-1}}}{\sqrt{1-\rho^{2(t_{il}-t_{il-1})}}}x^{**}_{it_{ij}}\nu_i) \\
&= \frac{\sqrt{1-\rho^2}}{n_i-1}\sum_{l=2}^{n_i}\frac{1-\rho^{t_{il}-t_{il-1}}}{\sqrt{1-\rho^{2(t_{il}-t_{il-1})}}}\mathbb{E}(x^{**}_{it_{ij}}\nu_i)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}(x^{**}_{it_{ij}}\overline{\overline{\nu^*}}) &= \frac{1}{\sum\limits_{l=1}^{N}(n_l-1)}\mathbb{E}(\sum_{l=1}^{N}\sum_{k=2}^{n_l} x^{**}_{it_{ij}}\nu^*_{lt_{lk}}) \\
&= \frac{1}{\sum\limits_{l=1}^{N}(n_l-1)}\mathbb{E}(\sum_{k=2}^{n_i} x^{**}_{it_{ij}}\nu^*_{it_{ik}}) \\
&= \frac{n_i-1}{\sum\limits_{l=1}^{N}(n_l-1)}\mathbb{E}(x^{**}_{it_{ij}}\overline{\nu^*_i}),
\end{aligned}
$$

the second equality following from the independency between $x^{**}_{it_{ij}}$ and $\nu^*_{lt_{lk}}$ for $l \neq i$, which yields $\mathbb{E}(x^{**}_{it_{ij}}\nu^*_{lt_{lk}}) = \mathbb{E}(x^{**}_{it_{ij}})\mathbb{E}(\nu^*_{lt_{lk}}) = 0$.

Since $\nu^{**}_{it_{ij}} = \nu^*_{it_{ij}} - \overline{\nu^*_i} + \overline{\overline{\nu^*}}$, we have:

$$
\begin{aligned}
&\mathbb{E}(x^{**}_{it_{ij}}\nu^{**}_{it_{ij}}) = \mathbb{E}(x^{**}_{it_{ij}}(\nu^*_{it_{ij}} - \overline{\nu^*_i} + \overline{\overline{\nu^*}})) \\
&= \sqrt{1-\rho^2}\frac{1-\rho^{t_{ij}-t_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}}\mathbb{E}(x^{**}_{it_{ij}}\nu_i) - (1-\frac{n_i-1}{\sum\limits_{i=1}^{N}(n_i-1)})\mathbb{E}(x^{**}_{it_{ij}}\overline{\nu^*_i}) \\
&= \sqrt{1-\rho^2}\frac{1-\rho^{t_{ij}-t_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}}\mathbb{E}(x^{**}_{it_{ij}}\nu_i) - (1-\frac{n_i-1}{\sum\limits_{i=1}^{N}(n_i-1)})\frac{\sqrt{1-\rho^2}}{n_i-1}\sum_{l=2}^{n_i}\frac{1-\rho^{t_{il}-t_{il-1}}}{\sqrt{1-\rho^{2(t_{il}-t_{il-1})}}}\mathbb{E}(x^{**}_{it_{ij}}\nu_i)
\end{aligned}
$$

Hence $\mathbb{E}(x^{**}_{it_{ij}}\nu^{**}_{it_{ij}}) = 0$ if and only if condition (20) holds or $\mathbb{E}(x^{**}_{it_{ij}}\nu_i) = 0$. This demonstrates Lemma 1.

We first consider equation (20). It writes:

$$\frac{1-\rho^{t_{ij}-t_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}} = (1 - \frac{n_i - 1}{\sum\limits_{i=1}^{N}(n_i - 1)})\frac{1}{n_i - 1}\sum_{l=2}^{n_i}\frac{1-\rho^{t_{il}-t_{il-1}}}{\sqrt{1-\rho^{2(t_{il}-t_{il-1})}}}$$

for all $2 \leq j \leq n_i$.

If these conditions were true, the quantities $\frac{1-\rho^{t_{ij}-t_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}}$, $2 \leq j \leq n_i$, would all be equal.
In the very particular case of the balanced case, all these conditions would be summed up in one single condition:

$$1 = 1 - \frac{1}{N}$$

It approximately holds for large N.
Hence, if we are in a balanced panel, these conditions are approximately respected.

Note that in the unbalanced case, if the data pattern were random, then condition (20) would be valid in expectancy (and for N large enough).

Lemma 2: Under the hypothesis that missing occurs independently of the observable $x_{i,t}$ and of the fixed-effects ν_i (i.e. the *missing-at-random* hypothesis), equation (20) holds in expectancy for N large enough.

We compute

$$\begin{aligned}
&\mathbb{E}((1-\frac{n_i-1}{\sum\limits_{i=1}^{N}(n_i-1)})\frac{1}{n_i-1}\sum_{l=2}^{n_i}\frac{1-\rho^{t_{il}-t_{il-1}}}{\sqrt{1-\rho^{2(t_{il}-t_{il-1})}}}|n_1,...,n_N)\\
&=(1-\frac{n_i-1}{\sum\limits_{i=1}^{N}(n_i-1)})\frac{1}{n_i-1}\sum_{l=2}^{n_i}\mathbb{E}(\frac{1-\rho^{t_{il}-t_{il-1}}}{\sqrt{1-\rho^{2(t_{il}-t_{il-1})}}}|n_1,...,n_N)\\
&=(1-\frac{n_i-1}{\sum\limits_{i=1}^{N}(n_i-1)})\frac{1}{n_i-1}(n_i-1)\mathbb{E}(\frac{1-\rho^{t_{ij}-t_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}}|n_1,...,n_N)\\
&=(1-\frac{n_i-1}{\sum\limits_{i=1}^{N}(n_i-1)})\mathbb{E}(\frac{1-\rho^{t_{ij}-t_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}}|n_1,...,n_N)\\
&=\mathbb{E}((1-\frac{n_i-1}{\sum\limits_{i=1}^{N}(n_i-1)})\frac{1-\rho^{t_{ij}-t_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}}|n_1,...,n_N)
\end{aligned}$$

the second equality following from the fair assumption that $\mathbb{E}(\frac{1-\rho^{t_{il}-t_{il-1}}}{\sqrt{1-\rho^{2(t_{il}-t_{il-1})}}}|n_i)$ does not depend on l. Taking expectations, we have

$$\begin{aligned}
&\mathbb{E}((1-\frac{n_i-1}{\sum\limits_{i=1}^{N}(n_i-1)})\frac{1}{n_i-1}\sum_{l=2}^{n_i}\frac{1-\rho^{t_{il}-t_{il-1}}}{\sqrt{1-\rho^{2(t_{il}-t_{il-1})}}})\\
&=\mathbb{E}((1-\frac{n_i-1}{\sum\limits_{i=1}^{N}(n_i-1)})\frac{1-\rho^{t_{ij}-t_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}})
\end{aligned}$$

Then, condition 20 holds in expectancy if and only if

$$\mathbb{E}(\frac{1-\rho^{t_{ij}-t_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}})=\mathbb{E}((1-\frac{n_i-1}{\sum\limits_{i=1}^{N}(n_i-1)})\frac{1-\rho^{t_{ij}-t_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}})$$

which is the case when N grows to infinity.

We focus here on the condition $\mathbb{E}(x^{**}_{it_{ij}}\nu_i) = 0$.
First, we compute

$$\mathbb{E}(x^{*}_{it_{ij}}\nu_i) = \frac{\sqrt{1-\rho^2}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}}\mathbb{E}((x_{it_{ij}} - \rho^{t_{il}-t_{il-1}}x_{it_{ij-1}})\nu_i)$$

$$\mathbb{E}(\overline{x^*_i}\nu_i) = \frac{\sqrt{1-\rho^2}}{n_i-1}\sum_{j=2}^{n_i}\mathbb{E}(\frac{x_{it_{ij}} - \rho^{t_{il}-t_{il-1}}x_{it_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}}\nu_i)$$

$$\mathbb{E}(\overline{\overline{x^*}}\nu_i) = \frac{n_i-1}{\sum\limits_{l=1}^{N}(n_l-1)}\mathbb{E}(\overline{x^*_i}\nu_i)$$

Hence

$$\begin{aligned}
&\mathbb{E}(x^{**}_{it_{ij}}\nu_i) = \mathbb{E}((x^{*}_{it_{ij}} - \overline{x^*_i} + \overline{\overline{x^*}})\nu_i)\\
&= \frac{\sqrt{1-\rho^2}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}}\mathbb{E}((x_{it_{ij}} - \rho^{t_{il}-t_{il-1}}x_{it_{ij-1}})\nu_i) - (1 - \frac{n_i-1}{\sum\limits_{l=1}^{N}(n_l-1)})\mathbb{E}(\overline{x^*_i}\nu_i)\\
&= \frac{\sqrt{1-\rho^2}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}}\mathbb{E}((x_{it_{ij}} - \rho^{t_{il}-t_{il-1}}x_{it_{ij-1}})\nu_i)\\
&- (1 - \frac{n_i-1}{\sum\limits_{l=1}^{N}(n_l-1)})\frac{\sqrt{1-\rho^2}}{n_i-1}\sum_{j=2}^{n_i}\mathbb{E}(\frac{x_{it_{ij}} - \rho^{t_{il}-t_{il-1}}x_{it_{ij-1}}}{\sqrt{1-\rho^{2(t_{ij}-t_{ij-1})}}}\nu_i)
\end{aligned}$$

A sufficient condition for $\mathbb{E}(x^{**}_{it_{ij}}\nu_i) = 0$ is that $\mathbb{E}(x^{*}_{it_{ij}}\nu_i)$ does not depend on j. Indeed, in this case,

$$\mathbb{E}(x^{**}_{it_{ij}}\nu_i) = \frac{n_i-1}{\sum\limits_{l=1}^{N}(n_l-1)}\mathbb{E}(x^{*}_{it_{ij}}\nu_i),$$

which converges to 0 when N tends to infinity.

Note however that the necessary hypothesis for this result ($(x^{*}_{it_{ij}}\nu_i)$ does not depend on j) depends on the value of ρ (through the Baltagi-and-Wu transformation) and is thus not necessarily very general.

B Proof of consistency of $\hat{\sigma_\epsilon}$

We have defined:

$$\widehat{\sigma_\varepsilon^2} = \frac{1}{N}\sum_{i=1}^{N}\overline{w}_i$$

where

$$\overline{w}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} w_{it_{ij}}$$

Noticing the variables $\overline{w}_i$'s are independent, we may apply the following version of the law of large numbers to $\overline{w}_i$:

Law of large number for independent non-identically distributed random variables: Let (X_i) be a sequence of independent random variables such that $\mathbb{E}[|X_i|^{1+\delta}] < \infty$ for some $\delta > 0$ and all i. Then, almost surely,

$$\frac{1}{N}\sum_{i=1}^{N} X_i - \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(X_i) \xrightarrow[N\to\infty]{} 0$$

This is enough to conclude as $\mathrm{E}(\overline{w}_i) = \mathrm{E}(w_{i,t}) = \sigma_\varepsilon^2$

C Comparison of the variance of two estimators of β

We consider the balanced case.

We denote e_1 and e_2 the error term after applying the within transformation to the first and second equation. We also denote with a tilde a variable whose individual mean has been substracted to. We have:

$$V(\hat{\beta}_1|X) = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'V(e_1|X)\tilde{X}(\tilde{X}'\tilde{X})^{-1}$$

$$V(\hat{\beta}_2|X) = (\tilde{X^{*2}}'\tilde{X^{*2}})^{-1}\tilde{X^{*2}}'V(e_2|X)\tilde{X^{*2}}(\tilde{X^{*2}}'\tilde{X^{*2}})^{-1}$$

If all terms of the second equation are multiplied by a constant a, a also multiplies e_2. And it has no impact on $V(\hat{\beta}_2|X)$. Hence, to ease the exposition, we can make all the computations as if our modification of the Baltagi-and-Wu transformation had been the following[14]:

$$\begin{aligned} x^{*2_{modified}}_{i,t_{i,j}} &= (1-\rho)x_{i,t_{i,j}} \text{ if } j = 1 \\ &= x_{i,t_{i,j}} - \rho x_{i,t_{i,j-1}} \text{ if } j > 1 \end{aligned} \tag{22}$$

We consider the diagonal terms of $V(e_2|X)$. It comes:

$$V(e_2|X)_{k,k} = (1 - \frac{1}{T})V(u^{*2_{modified}}|X)_{k,k}$$

$$\begin{aligned} V(e_2|X)_{k,k} &= (1-\frac{1}{T})V((1-\rho)u|X)_{k,k} = (1-\frac{1}{T})\frac{(1-\rho)^2}{(1-\rho^2)}\sigma^2_\varepsilon \text{ if } j = 1 \\ &= (1-\frac{1}{T})V(\varepsilon|X)_{k,k} = (1-\frac{1}{T})\sigma^2_\varepsilon \text{ if } j > 1 \end{aligned} \tag{23}$$

We consider the off-diagonal terms of $V(e_2|X)$. They are null when considering covariances that are not about the same individual. Otherwise, they are equal to: $-\frac{2}{T}(1-\frac{1}{T})\sigma^2_\epsilon$ when they do not concern the first observation of an individual.

When making the same computations, we get similar results except that: $V(\hat{\beta}_1|X) =$

[14] We denote $a = \frac{1-\rho}{\sqrt{1-\rho^2}}$

$(X'X)^{-1}X'V(e|X)X(X'X)^{-1}$.
Diagonal terms of $V(e|X)$ are:

$$\begin{aligned}\mathrm{Var}(e_{1,it}) &= \mathrm{E}((u_{it}-\overline{u}_i)^2)\\ &= \sigma_u^2 + \frac{1}{T^2}\sum_{p,s=1}^{T}\rho^{|p-s|}\sigma_u^2 - \frac{2}{T}\sum_{p=1}^{T}\rho^{|p-t|}\sigma_u^2\end{aligned}$$

One notices (as demonstrated by Appendix B of Bhargava *et al.* (1982)):

$$\sum_{j,k=1}^{T}\rho^{|j-k|} = \frac{1+\rho}{1-\rho}T - \frac{2\rho}{1-\rho}\frac{1-\rho^T}{1-\rho}$$

And:

$$\sum_{p=1}^{T}\rho^{|p-t|} = \frac{1-\rho^t}{1-\rho} + \rho\frac{1-\rho^{T-t}}{1-\rho}$$

And non-diagonal terms are:

$$\begin{aligned}\mathrm{Cov}(e_{1,it}, e_{1,is}) &= \mathrm{Cov}(u_{it}-\overline{u}_i, u_{is}-\overline{u}_i)\\ &= \rho^{|t-s|}\sigma_u^2 - \frac{1}{T}\sum_{p=1}^{T}\rho^{|p-t|}\sigma_u^2 - \frac{1}{T}\sum_{p=1}^{T}\rho^{|p-s|}\sigma_u^2 + \frac{1}{T^2}\sum_{p,s=1}^{T}\rho^{|p-s|}\sigma_u^2\end{aligned}$$

Let us suppose here that T is large enough so that all terms of order $\frac{1}{T}$ can be neglected. In this case:
$V(e_2)$ is close to a diagonal matrix whose diagonal terms are σ_ϵ^2.[15] $V(e_1)$ is close to a matrix whose cross-individual terms are nul, but whose terms over individuals are: $\rho^{p-s}\frac{\sigma_\epsilon^2}{1-\rho^2}$.

[15] Terms regarding the first observations of an individual are slightly different. For instance the corresponding diagonal terms are close to $\frac{(1-\rho)^2}{1-\rho^2}\sigma_\epsilon^2 < \sigma_\epsilon^2$