Monte Carlo experiment on the pooled ols estimator in large mixed panels.

marchese, malvina

London School of Economics Political Science

2009

MONTE CARLO EXPERIMENTS ON THE OLS ESTIMATOR IN PANEL DATA WITH STATIONARY AND NONSTATIONARY REGRESSORS

Abstract.

This paper investigates the behaviour of the pooled ols estimator in a panel data model with stationary and nonstationary regressors as both n and T go to infinity.The nonstationary regressor is assumed I(1), the stationary regressor is set i.i.d.The investigation is carried through four Monte Carlo experiments.The experiments show that in a model with no endogeneity,the pooled ols estimator of the I(1) regressor is $\sqrt{n}T$ consistent and asymptotically normal and the estimator of the I(0) regressor is $\sqrt{nT}$ consistent and asymptotically normal.If a correlation between the I(1) regressor and the regression disturbance is introduced, both estimators are inconsistent for small n but $\sqrt{nT}$ consistent and asymptotically normal for large n and T.When an individual random effect is added to the basic model the results for the I(0) regressor coefficient do not alter,whereas the estimator of the I(1) regressor loses in efficiency and becomes $\sqrt{nT}$ consistent and asymptotically normal.

1.Introduction

The advantages of panel data sets over cross-sectional and time series data have long been established in econometric research.

Panel data sets usually give the researcher a larger number of data points than conventional cross-section and time series data, thus increasing the degrees of freedom and reducing collinearity among explanatory variables. This results in more reliable parameter estimates and,most importantly,enables the researcher to specify and test more sophisticated models with less restrictive behavioural assumptions.These data sets make it possible to identify and measure effects that are simply not detectable in pure cross-section or time series data and to eliminate or reduce the estimation bias.

The initial focus of panel data research has been on identifying and estimating effects form stationary panels with a large number of cross section data (n) and few time series observations (T).A limiting theory for the pooled ordinary least squares estimator(pooled ols) in this setting has been well established since the works of Hsiao(1986) and Chamberlain(1984).

However more recently empirical work in econometrics has used panel data for which the time series component is nonstationary with both large n and T available.Examples of this literature range from testing growth convergence theories in macroeconomic to estimating long run relations between international financial series such as relative prices and exchange rates. These works have been facilitated and enhanced by the availability of a number of important panel data sets covering different individuals,regions and countries over a relatively long period of time, for example the Penn World table. Nonstationary panels provide a further instance of the ability of panel data to identify effects that time series or cross section data alone cannot identify.

When the time series component of the model is allowed to be non stationary and both large n and T are taken into account the traditional limiting theory for the pooled ols estimator is no longer valid.Phillips and Moon (1999) investigated regressions with non stationary panel data for which the time series component is an integrated of order one process, I(1) and where both n and T are large.They have shown that, under a variety of different cointegrating relations between the regressors and the regressand, the pooled ols estimator is consistent and asymptotically normal.

The limiting theory developed by Phillips and Moon for non stationary panels does not allow for the presence of both stationary and non stationary regressors in the same model.In practise however this framework is very relevant.In the economic literature there are many cases of models with data that are mixed stationary weekly dependent ,I(0), and unit root processes, I(1).A quite important one is the analysis of demand systems where budget shares or quantities are regressed on relative prices and real income for different countries over time. Typically some relative prices are quite stable, I(0), and some other are trending I(1).Money demand equations offer a similar mixture of stationary and nonstationary variables, with real income trending over time for most countries but stationary interest rate.

A comprehensive limiting distribution theory for the pooled ols estimator in this framework has not yet been developed.Very recently Baltagi,Kao and Liu (2008) have developed a limit theory for the pooled ols estimator in a simple panel regression model with random error component disturbances.They assume that both the regressor and the remainder disturbance term are autoregressive and possibly non stationary and derive asymptotic distribution results for the pooled ols estimator when $T \rightarrow \infty$ followed by $n \rightarrow \infty$.They show that in a model with random error components the ols estimator has a normal asymptotic distribution and different rate of convergence dependent on the non stationarity of the regressors and of the remainder disturbance.When the disturbance term is assumed I(0) and the regressor I(1) or viceversa.The limit theory they develop is specific for random error component disturbance models. While the random error component model is the most largely used in micro panel data research, a more general framework is required for macro panel data models.The results of Baltagi, Kao and Liu cannot be extended to a different panel structure, thus further investigation into the asymptotic behaviour of the pooled ols estimator in a less specific panel model is desirable.

The aim of this paper is to develop such an investigation.The paper wants to provide insight into the behaviour of the pooled ols estimator when n and T are allowed to go to infinity simultaneously a very general panel data model with both integrated of order one and stationary regressors.Results of a Monte Carlo experiment on the consistency and asymptotic normality of the ols estimator in this setting are presented. I then investigate the behaviour of the estimator under circumstances that in traditional panel data limit theory would generate inconsistency in the estimates. Two potential sources of inconsistency for the ols estimator are separately added to the model by introducing correlation between the regression error and, respectively, the stationary and nonstationary regressor. Interesting enough when the potential source of inconsistency is introduced trough a correlation between the regression error and the I(1) regressor, I find that the pooled ols estimator is still consistent as T is allowed to go to infinity.Finally I extend the model to allow for individual heterogeneity in a random effect framework, with no correlation between the regression error and the individual effect. Individual heterogeneity is introduced allowing the intercept of the model to vary across individuals. The ols estimator is found to be consistent and asymptotically normal but with some losing in efficiency with respect to the initial case.

The structure of the paper is as follows.Section 2 presents a brief overview of the literature on non stationary panel data model with large n and T. Section 3 lays out the basic model and assumptions and presents Monte Carlo simulation results on the behaviour of the ols estimators of the I(1) and I(0) regressors.Section 4 gives insight into the behaviour of the estimator when the regression error and one of the regressors are correlated. Section 5 discuss the individual effect model .Section 6 draws conclusions.

2.Literature review.

Since the beginning of the 1990's there has been much ongoing research on nonstationary panel data.

Quah(1994), Levin and Lin(1993) consider unit root time series regressions with non stationary panel data and propose test statistics for unit roots .Pedroni(1995) studies some properties of cointegration statistics in pooled time series panels,Robertson and Symons(1992)study the bias that are likely to arise in practise with nonstationary panel data .Baltagi and Kramer(1997) and, more recently Kao and Emerson(2004) investigate the case of a panel time trend model.

Pesaran and Smith (1995) examine the impact of nonstationary variables on cross section regression estimates with a large number of groups (n) available and a large number of time periods(T).Assuming that the parameter of interest is the average effect of some exogenous variable on a dependent variable, they argue that when T is large enough it is sensible to run separate regression for each cross-section group. In particular they examine the impact of nonstationary variables on the cross-section estimates. Under some quite strong assumptions such as exogeneity of the regressors and iid disturbances,they show that no spurious correlation will arise between two I(1) variables and that the cross-section ols estimator of the average effect will be consistent for large T.

Phillips and Moon (1999) extend the work of Pesaran and Smith to a very general setting and present a fundamental framework for studying asymptotic behaviour of the ols estimator in non stationary panel data models with large n and T.Their work investigates the behaviour of the pooled ols estimator in panel data models where all the regressors are non stationary under four possible panel structure: when there is no cointegrating relation between regressors and regressand, when there is a heterogeneous cointegration relation, when the cointegrating relation is homogenous and near-homogenous.

When there is no cointegrating relation between the regressors and the regressand, the model is know in the literature as spurious regression. If panel observations with both large cross-sectional and time series components are available,then, even if the noise is quite strong,it can be characterised as independent across individuals. By pooling the cross section and the time series observations, Phillips and Moon attenuate the strong effect of the residuals in the regression while retaining the strength of the signal . They show the existence of a very interesting long run average relationship between the regressors and the regressand and they prove that the pooled ols estimator of such relation is both $\sqrt{n}$consistent and asymptotically normal.

When the existence of a cointegrating relations between regressors and regressand is assumed across all the individuals,the limiting distribution of the pooled ols estimator is derived for three cases:heterogenous, homogenous and near homogenous cointegration. A cointegrating relation between the regressors and the regressand exists when their conditional long run variance matrix has deficient rank. If different cointegrating relationships are allowed across individuals the model is know as a heterogeneous cointegration model. When the cointegration relation is the same for all individuals the model is said to be a homogenous cointegration model,if only slightly different conintegrating relations exist across individuals the model is a near homogenous panel cointegration model. In the first case Phillips and Moon show that the pooled ols estimator consistently estimates the long run average coefficient between regressors and regressand. By the same logic of the spurious regression model,the consistency is obtained because, by pooling the panel data,cross section pooling attenuates the strength of the noise relative to the signal of the regression. The ols estimator is found to be $\sqrt{n}$ consistent and asymptotically normal. The same estimator is proved n$\sqrt{T}$ consistent and asymptotically normal both in the homogenous and in the near homogenous cointegration model.

The development of a limiting distribution theory in panel data with both large n and T requires deriving the limiting behaviour of double indexed processes. In general this limiting behaviour depends on the treatment of the two indexes,n and T. Different approaches are available. One approach is to fix one index,n ,and allow the other,T,to pass to infinity,giving an intermediate limit. By letting n pass to infinity subsequently a sequential limit theory is obtained. A second approach is to let the indexes pass to infinity along a specific diagonal path determined by a monotonically increasing function relation of the type T=T(N) where n$\rightarrow \infty$. This approach is known as diagonal path limit theory. A third approach is to allow both indexes to pass to infinity simultaneously without pacing any restriction on the path of divergence. This approach is known as joint limit theory.Diagonal path limit theory requires assuming a very specific expansion path and thus may fail to provide an appropriate approximation for a given (N,T) situation. Joint limit theory requires stronger conditions than sequential one but ,on the other hand,sequential limits can give asymptotic results that are misleading in cases where both indexes pass to infinity simultaneously.

The limit theory developed by Phillips and Moon allows for both sequential and joint limits. To derive the latter they impose the rate condition n\T $\rightarrow$ 0,thus in practise their limit theory is most likely to be useful when n is moderate and T is large. Such data configuration can be expected in multi-country macroeconomic data,for example, when attention is restricted to group of countries such as OECD nations or developing countries. The results stated above on the asymptotic normality and consistency of the ols estimator are derived ,in all four panel structure, using both sequential limit theory and, under some strengthening of the conditions, joint limit theory as well.

The work of Phillips and Moon provides a very exhaustive investigation about the limiting behaviour of the pooled ols estimator panel data with large n and T when all the regressors are assumed to be nonstationary, I(1).At present the only work in the literature about the asymptotics of the ols estimator in panel data model with large n and T and with mixed I(1) and I(0) regressors is Baltagi,Kao,Liu(2008).This paper studies the asymptotic properties of the ols estimator in a simple panel data model with random error component disturbance. Both the regressor and the remainder disturbance term are assumed to be autoregressive and possibly non stationary.Baltagi, Kao and Liu present results for the case of one regressor, but their results could easily be extended to the multiple regressor case. They consider the following model:

$$y_i = \alpha + x_{it}\beta + u_{it}$$

with i=1,....,n and t=1,......T and where $u_{it}=\mu_i + \nu_{it}$ and α and β are scalars. The individual effect μ_i is assumed to be random and i.i.d

$N(0,\sigma^2)$.$\{\nu_{it}\}$ is set to be an AR(1) process:

$$\nu_{it} = \rho\nu_{it} + \varepsilon_{it}$$

with $|\rho| \prec 1$ and ε_{it} white noise process. The μ_i are assumed independent of the ν_{it} for all i and T.This is the random effects error component model with serial correlation.They further assume that

$$x_{it} = \lambda x_{it-1} + \epsilon_{it}$$

with $|\lambda| \prec 1$ and ϵ_{it} white noise process, and that $E(\mu_i|x_{it})=0$ for all i and T (the usual fixed effect model assumption).

The pooled ols estimator of β is given by:

$$\hat{\beta}_{ols} = \frac{\sum_i \sum_t (x_{it} - \tilde{x})(y_{it} - \tilde{y})}{\sum_i \sum_t (x_{it} - \tilde{x})}$$

where $\tilde{x}=\frac{1}{nT}\sum_i \sum_t x_{it}$ and $\tilde{y}=\frac{1}{nT}\sum_i \sum_t y_{it}$.

All the asymptotic results of Baltagi, Kao and Liu are based on a sequential limit theory, in particular they assume that $T \rightarrow \infty$ followed by $n \rightarrow \infty$.They do not provide any joint limit distributional results .Their work shows that the asymptotic properties of the ols estimator depend crucially on the serial correlation properties of the regressor and the error component in the disturbance term (ν_{it}).When the regression error and the regressor are both stationary ($|\rho| \prec 1$ and $|\lambda| \prec 1$) the ols estimator is found to be$\sqrt{nT}$ consistent and asymptotically normal.If the disturbance and is I(1) and the regressor is I(0) ($\rho = 1$ and $|\lambda| \prec 1$) the ols estimator is$\sqrt{n}$consistent and asymptotically normal. The noise is so strong that it dominates the signal. It is interesting to notice that, when only few time series observations are available for this case ,the ols estimator is inconsistent.When the disturbance is I(0) and the regressor is I(1) ($|\rho| \prec 1$ and $\lambda = 1$) the model is cointegrated:the ols estimator is $\sqrt{nT}$consistent and asymptotically normalIf both the disturbance and the regressor are I(1) ($\rho = 1$ and $\lambda = 1$) the regression is spurious and the estimator is $\sqrt{n}$consistent and asymptotically normal.

3. A panel model with stationary and nonstationary regressors.

In this section I extend the model of Baltagi,Kao and Liu to a more general panel data setting and present results from a Monte Carlo simulation to gain insight into the behaviour of the pooled ols estimator in this model.The model I consider is the following:

$$y_{it} = \alpha + \beta x_{it} + \gamma z_{it} + u_{it}$$

where i=1,.....,n and T=1,......,T.In the model α, β and γ are scalar, however the results can be easily extended to the case of vector regressors.The regression disturbance u_{it} is i.i.d ,z_{it} is a stationary weakly dependent process(e.g. i.id) and x_{it} is a unit root process:

$$x_{it} = x_{it-1} + \eta_{it}$$

where η_{it} is white noise, independent from u_{it}.

The pooled ols estimator of the model is given by

$$\hat{\vartheta} = \frac{\sum_i \sum_t x_{it} y_{it}}{\sum_i \sum_t x_{it}^2}$$

.

$\hat{\vartheta}_{it}$ is thus a(3*1) vector of estimates whose first and second components are respectively the pooled estimator of the coefficient of the I(1) regressor $(\hat{\beta})$ and the pooled ols estimator of the coefficient of the I(0) regressor $(\hat{\gamma})$.

In the Monte Carlo experiment the true value of $\vartheta=(\alpha,\beta,\gamma)'$ is set at (1,1,1)'. z_{it}, x_{it} and η_{it} are generated as independent Normal(0,1). The values of n and T range from 10 to 500, and for each possible combination of n and T the reported results are the outcome of 10000 simulations.

The range of values of T and n and the number of simulations have been set according to the common practise of Monte Carlo simulation for panel data models.

In this setting any inconsistency of the estimators can be clearly detected within the set range of variation of T and n.

For consistency of the estimators a sufficient and, almost close to necessary, condition is the decrease of the standard deviation and the mean of the difference between the true and the simulated value of the estimators. When both std($\hat{\beta}-\beta$) and E($\hat{\beta}-\beta$) decrease as n and T increase, the evidence from the experiment is in favour of consistency.

Results for the std($\hat{\beta}-\beta$) and E($\hat{\beta}-\beta$) are reported below:

T

		10	100	200	300	500
	10	0.00485	0.0028	0.0024	0.0013	9.3891×10^{-4}
n	100	0.0136	0.0014	7.1200×10^{-4}	4.4321×10^{-4}	2.8198×10^{-4}
	200	0.0096	9.9293×10^{-4}	5.0075×10^{-4}	3.8790×10^{-4}	1.9894×10^{-4}
	300	0.0052	8.4368×10^{-4}	3.2165×10^{-4}	2.0813×10^{-4}	1.5168×10^{-4}
	500	0.0061	6.2840×10^{-4}	3.1422×10^{-4}	2.0093×10^{-4}	1.2632×10^{-4}

$Std(\hat{\beta}-\beta)$

T

		10	100	200	300	500
	10	0.0357	0.0038	0.0019	0.0008	7.3626×10^{-4}
n	100	0.0108	0.0011	5.6657×10^{-4}	4.0351×10^{-4}	2.2446×10^{-4}
	200	0.0077	7.9293×10^{-4}	3.9931×10^{-4}	2.6139×10^{-4}	1.5855×10^{-4}
	300	0.0059	6.9821×10^{-4}	3.0803×10^{-4}	2.5444×10^{-4}	1.4982×10^{-4}
	500	0.0048	5.0199×10^{-4}	2.5133×10^{-4}	2.0075×10^{-4}	1.0105×10^{-4}

$E(\hat{\beta}-\beta)$

From the reported results there are evidences to argue for consistency of the estimator of coefficient of the I(1) regressor.

When T is small and fixed, say T=10, and n is allowed to go to infinity the evidence of consistency is coherent with the theoretical results developed by Phillips(1986) and Granger and Newbold(1974).

Furthermore the experiment brings evidences for consistency of the estimator for the case of small, fixed n as T is allowed to go to infinity; and for the case where both n and T go to infinity simultaneously.

To assess the rate of consistency of the estimator I regress the logarithm of the variance of the simulated estimator on the logarithm of n and of T.The regression equation is the following:

$$\log(Var(\hat{\beta})) = \log(C) - \phi\log(n) - \psi\log(T)$$

where C is the constant term.

After discarding the values of the variance of $\hat{\beta}$ for all the cases where n=10, and for all the cases where T=10 the following least squares estimates of φ and ψ are obtained

$$\begin{aligned}
\hat{\phi} &= 1.0310 \\
\hat{\psi} &= 1.9919 \\
R^2 &= 0.9998 \\
pvalue &= 0.000 \\
Fstatistic &= 62.2499 \\
\hat{\sigma}^2 &= 0.5458
\end{aligned}$$

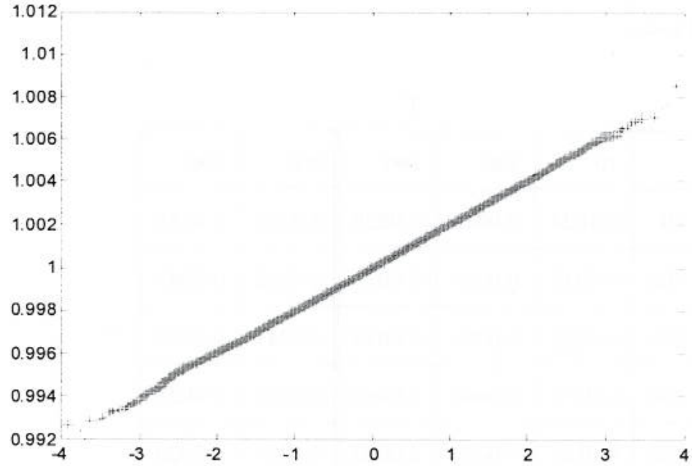

Figure 1: Q-Q plot $\hat{\beta}$ experiment one

For asymptotic normality I make use of a very common graphical statistic device: a Q-Q plot.A Q-Q plot is the most widely used graphical method for diagnosing difference between the probability distribution of the statistic of interest and a normal distribution. The plot on top of the page is the Q-Q plot of $\hat{\beta}$ for the case of (n,T)=(500,500).The sample size is of 10000 observations. The plot reports the quantile of the normal distribution on the horizontal axis and the order statistics of the simulated sample on the vertical axis.When the distribution of the simulated sample is the same of the comparison normal distribution the plot approximates a straight line.

From the above results the experiment brings evidences for $\sqrt{n}$T consistency and asymptotic normality of β,the coefficient of the I(1) regressor.

Results for the standard deviation of $(\hat{\gamma} - \gamma)$ and for the $E(\hat{\gamma} - \gamma)$ are reported below:

T

		10	100	200	300	500
	10	0.1024	0.0313	0.0223	0.0201	0.0141
n	100	0.0315	0.0099	0.0071	0.0063	0.0045
	200	0.0222	0.0071	0.0049	0.0041	0.0032
	300	0.0217	0.0065	0.0042	0.0033	0.0027
	500	0.0142	0.0045	0.0032	0.0028	0.0020

$Std(\hat{\gamma}-\gamma)$

T

		10	100	200	300	500
	10	0.0816	0.0250	0.0179	0.0161	0.0112
n	100	0.0251	0.0078	0.0056	0.0046	0.0036
	200	0.0177	0.0057	0.0040	0.0036	0.0025
	300	0.0141	0.0048	0.0031	0.0027	0.0023
	500	0.0113	0.0035	0.0025	0.0021	0.0016

$E(\hat{\gamma}-\gamma)$

The least square estimates of φ and ψ in the regression of the logarithm of the variance of $\hat{\gamma}$on the logarithm of n and the logarithm of T are the following:

$$\begin{aligned}
\hat{\varphi} &= 1.0053 \\
\hat{\psi} &= 1.0059 \\
R^2 &= 0.9998 \\
pvalue &= 0.000 \\
Fstatistic &= 62.2499 \\
\hat{\sigma}^2 &= 0.5458
\end{aligned}$$

The Q-Q plot of $\hat{\gamma}$ against the Normal distribution for (n,T)=(500,500) is reported on top of next page.

According the evidences presented above the coefficient of the stationary regressor,γ. is $\sqrt{nT}$consistent and asymptotically normal.

4.Introducing possible sources of inconsistency.

All the above results have been derived assuming that:

1) the regression error u_{it} and the exogenous shock of the AR(1) process generating the non stationary regressor, η_{it}, are independent;

2) the regression error u_{it} and the I(0) regressor z_{it} are independent

According to the conventional limit theory of the pooled ols estimator of panel models,any correlation between the regressors and the regression disturbance will result in an inconsistency of the estimator.However these conclusions

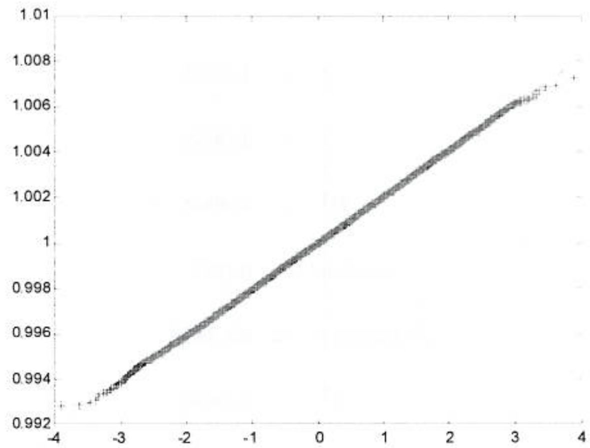

Figure 2: Q-Q plot $\hat{\gamma}$ experiment one

are derived under the assumption of a small number of time series observations,T, as the number of cross section observations ,n ,goes to infinity.

In other settings than the one under consideration it has already be shown that when the number of time series observations is allowed to go to infinity, the conventional results of inconsistency may no longer hold.Baltagi,Kao and Liu(2008) have proven that in a model with random error component disturbances when the regressor is I(0) and the disturbance is I(1) the ols estimator is still consistent if large and and T are available, whereas the traditional result of inconsistency continues to hold if only few time series observations are available.

It thus appears quite relevant to investigate the behaviour of the ols estimator when some correlation between the regressors and the regression disturbance is introduced in the basic model.

Recall the model:

$$y = \alpha + \beta x_{it} + \gamma z_{it} + u_{it}$$

for i=1,......n and t=1........T and

$$x_{it} = x_{it-1} + \eta_{it}$$

where η_{it} and u_{it} are generated as independent Normal(0,1) but z_{it} is generated as

$$z_{it} = \rho u_{it} + \chi\sqrt{1-\rho}$$

where $\chi \sim N(0,1)$. A correlation of ρ is now introduced between the regression disturbance and the I(0) regressor.

The tables below show the results for the E($\hat{\beta}-\beta$) and E($\hat{\gamma}-\gamma$) for $\rho = 0.2$.

The results are robust to higher values of ρ .Both the E($\hat{\beta}-\beta$) and the E($\hat{\gamma}-\gamma$) fail to decrease along all the dimensions of the experiment. There are clear evidences of inconsistency.

T

	10	100	200	300	500
	0.1210	0.0131	0.0130	0.0132	0.0131
100	0.1193	0.0129	0.0128	0.0133	0.0133
n 200	0.1196	0.0134	0.0134	0.0134	0.0134
300	0.1198	0.0132	0.0129	0.0131	0.0131
500	0.1198	0.0133	0.0133	0.0132	0.0132

$E(\hat{\beta}-\beta)$

T

	10	100	200	300	500
10	0.4952	0.4953	0.4962	0.4977	0.4991
n 100	0.4551	0.4952	0.4975	0.4984	0.4991
200	0.4552	0.4952	0.4976	0.4985	0.4992
300	0.4554	0.4953	0.0976	0.4984	0.4991
500	0.4554	0.4952	0.4976	0.4984	0.4990

$E(\hat{\gamma}-\gamma)$

The results of the experiment are quite different when the potential source of inconsistency is introduced trough a correlation between the regression error and the I(1) regressor.In the above model the regression error u_{it} and the I(0) regressor x_{it} are now generated as independent Normal(0,1).But η_{it} is now generated as:

$$\eta_{it} = \rho u_{it} + \chi\sqrt{1-\rho}$$

where $\chi \sim N(0,1)$. A correlation of ρ is now introduced between the regression disturbance and the shock of the AR(1) process generating the nonstationary regressor.Clearly this brings in the model a correlation between the error and the I(1) regressor.

Results for the standard deviation of ($\hat{\beta}-\beta$) and for the E($\hat{\beta}-\beta$) for $\rho = 0.8$ .The results are robust to lower values of ρ.

T

n		10	100	200	300	500
	10	0.0411	0.0049	0.0022	0.0015	9.1871×10^{-4}
n	100	0.0112	0.0009	6.1787×10^{-4}	4.1625×10^{-4}	2.4681×10^{-4}
	200	0.0078	8.6225e-004	4.3374×10^{-4}	2.9828×10^{-4}	1.7634×10^{-4}
	300	0.0064	7.0626e-004	3.9421×10^{-4}	2.3663×10^{-4}	1.4067×10^{-4}
	500	0.0056	5.4675e-004	2.7225×10^{-4}	1.1811×10^{-4}	1.1029×10^{-4}

$std(\hat{\beta}-\beta)$

T

n		10	100	200	300	500
	10	0.1210	0.0131	0.0066	0.0043	0.0026
n	100	0.1130	0.0120	0.0060	0.0040	0.0024
	200	0.1128	0.0120	0.0060	0.0040	0.0024
	300	0.1125	0.0119	0.0060	0.0040	0.0024
	500	0.1125	0.0119	0.0060	0.0040	0.0024

$E(\hat{\beta}-\beta)$

Not surprisingly the standard deviation of $\hat{\beta}-\beta$ decreases as the number of observations increases along any dimension.However the reported results show an interesting behaviour of the bias in the experiments. When T is fixed at any value of the range and n is allowed to go to infinity, the estimator is inconsistent as we would expect from the conventional limit theory.But when n is fixed at any value of the range ,and T is allowed to go to infinity the bias decreases: there are evidences for consistency of the estimator.Analogously if both n and T are allowed to go to infinity simultaneously,providing thus a sort of experimental evidence for a joint limit theory,the bias decreases. The evidences support consistency.When a regression of the log of the variance of the estimator on the log of n and T is run the following estimates are obtained:

$$\begin{aligned} \hat{\varphi} &= 0.9183 \\ \hat{\psi} &= 0.8921 \\ R^2 &= 0.8907 \\ pvalue &= 0.000 \\ Fstatistics &= 60.0033 \\ \hat{\sigma}^2 &= 0.5998 \end{aligned}$$

For the asymptotic normality of the estimator as n and T go to infinity simultaneously the Q-Q plot on top of next page is obtained for (n,T) =(500,500).

Interesting enough the experiment brings evidences for $\sqrt{nT}$ consistency and asymptotic normality of the ols estimator of the coefficient of the I(1) regressor

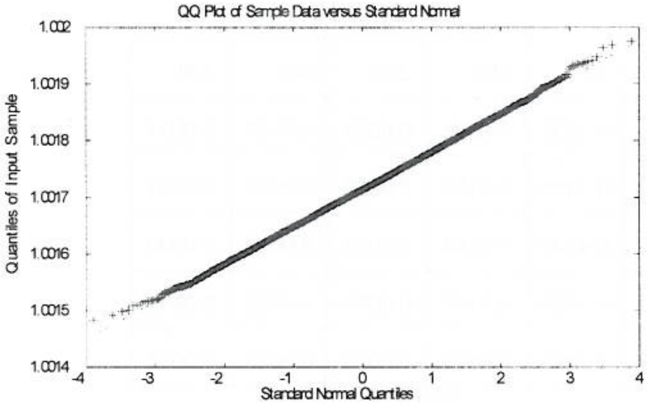

Figure 3: Q-Q plot $\hat{\beta}$ experiment 3

when the number of cross section observations and the number of time series observations are allowed to go to infinity simultaneously.

The behaviour of the estimator of the coefficient of the I(0) regressor is analogous.Results for the standard deviation of ($\hat{\gamma}-\gamma$) and for the E($\hat{\gamma}-\gamma$) are reported below:

T

		10	100	200	300	500
	10	0.0448	0.0051	0.0026	0.0017	0.0011
n	100	0.0125	0.0014	7.0786×10^{-4}	4.7339×10^{-4}	2.9213×10^{-4}
	200	0.0086	7.9162×10^{-4}	4.9539×10^{-4}	3.2350×10^{-4}	2.8812×10^{-4}
	300	0.0070	6.9226×10^{-4}	4.0631×10^{-4}	2.9398×10^{-4}	1.6378×10^{-4}
	500	0.0057	2.9815×10^{-4}	2.0761×10^{-4}	1.9864×10^{-4}	1.3004×10^{-4}

$Std(\hat{\gamma}-\gamma)$

		T				
		10	100	200	300	500
	10	0.1531	0.0168	0.0085	0.0057	0.0034
n	100	0.1436	0.0156	0.0079	0.0052	0.0031
	200	0.4130	0.0154	0.0078	0.0052	0.0030
	300	0.1429	0.0156	0.0079	0.0052	0.0031
	500	0.1429	0.0156	0.0078	0.0052	0.0031

$E(\hat{\gamma}-\gamma)$

For the rate of consistency the following estimates are obtained:

$$
\begin{aligned}
\hat{\varphi} &= 0.9999 \\
\hat{\psi} &= 0.9981 \\
R^2 &= 0.8907 \\
pvalue &= 0.000 \\
Fstatistics &= 60.0033 \\
\hat{\sigma}^2 &= 0.5998
\end{aligned}
$$

The asymptotic normality is supported by the Q-Q plot of $\hat{\gamma}$ for (n,T)=(500,500) on top of next page.

The experiment brings evidences for $\sqrt{nT}$ consistency and asymptotic normality of the ols estimator of the stationary coefficient.

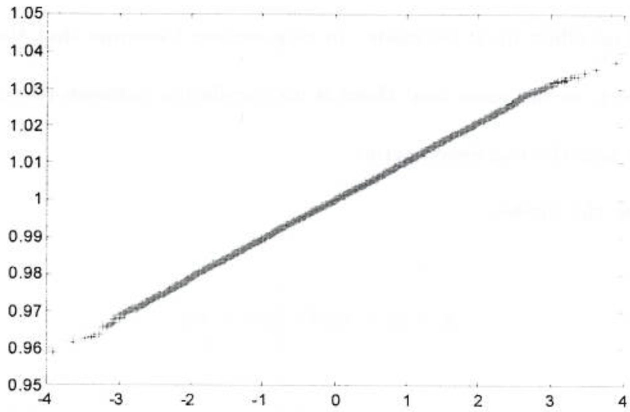

Figure 4: Q-Q plot $\hat{\gamma}$ experiment 3

5.Introducing individual effect.

One of the advantages of panel data sets is that they enable the researcher to allow for the presence of individual heterogeneity in the model.There is no single specification of individual heterogeneity which is universally valid:in fact the choice of the appropriate specification depends on the problem on hand and on the nature of the data.In macroeconomic panel data models it is quite a common practise to assume that the reaction coefficients are the same for all individuals and to account for individual heterogeneity by allowing a different intercept across individuals.In this section I introduce an individual effect in the basic model by allowing for different intercepts across individuals .I then investigate the behaviour of the pooled ols estimator once this extra source of generality is allowed. As soon as individual heterogeneity, in any form, is introduced in the model, an assumption on the correlation between the regression error and the individual effect must be made. In this section I assume that the model is random effect, in the sense that there is no correlation between the unobserved component and the regression error.

Consider the model:

$$y = \alpha_i + x_{it}\beta + z_{it}\gamma + u_{it}$$

for i=1,..........,n and t=1,.........,T. Where z_{it} is a stationary I(0) regressor generated as a N(0,1) and independent of u_{it}. $\alpha_i$is an individual effect generated as a N(0,1) independent of u_{it} . And the I(1) regressor is

$$x_{it} = x_{it-1} + \eta_{it}$$

where η_{it} is generated as a Normal(0,1) independent from u_{it} and from α_{it}.

In the experiment the range of n and T and the number of simulations have been chosen as in the previous two.

The tables below reports results for the standard deviation of $(\hat{\beta} - \beta)$ and for $E(\hat{\beta} - \beta)$

T

n		10	100	200	300	500
	10	0.1148	0.0352	0.0251	0.0211	0.0157
n	100	0.0359	0.0112	0.0079	0.0050	0.0033
	200	0.0253	0.0079	0.0056	0.0046	0.0028
	300	0.0219	0.0061	0.0043	0.0035	0.0021
	500	0.0160	0.0050	0.0035	0.0029	0.0012

$std(\hat{\beta}-\beta)$

T

n		10	100	200	300	500
	10	0.0914	0.0281	0.0251	0.0211	0.0157
n	100	0.0286	0.0089	0.0063	0.0052	0.0044
	200	0.0202	0.0063	0.0044	0.0041	0.0033
	300	0.0167	0.0058	0.0054	0.0039	0.0022
	500	0.0128	0.0040	0.0028	0.0023	0.0019

$E(\hat{\beta}-\beta)$

Both statistics decrease along all the dimensions of the experiment: there are evidences for consistency. For the rate of convergence the following estimates are obtained

$$\begin{aligned}
\hat{\varphi} &= 0.9983 \\
\hat{\psi} &= 1.0341 \\
R^2 &= 0.9154 \\
pvalue &= 0.000 \\
Fstatistic &= 63.4478 \\
\hat{\sigma}^2 &= 0.4915
\end{aligned}$$

The Q-Q plot of $\hat{\beta}$ for (n,T)=(500,500) reported on top of next page gives evidences for asymptotic normality.

On the ground of the above results it appears that in this case the ols estimator of the coefficient of the I(1) regressor is $\sqrt{nT}$ consistent and asymptotically normal.

When comparing this results with the first case presented in the paper it is evident that the introduction of an individual effect has determined a loss in efficiency in the estimator of the I(1). This result is coherent with the founding of Baltagi, Kao and Liu for the case of a I(1) regressor and of a I(0) remainder disturbance. However in the random error component model they investigate the I(0) component is introduced as the remainder disturbance and thus has no coefficient to estimate. In the model presented above the I(0) component is introduced as one of the regressors, thus further investigation into the asymptotic

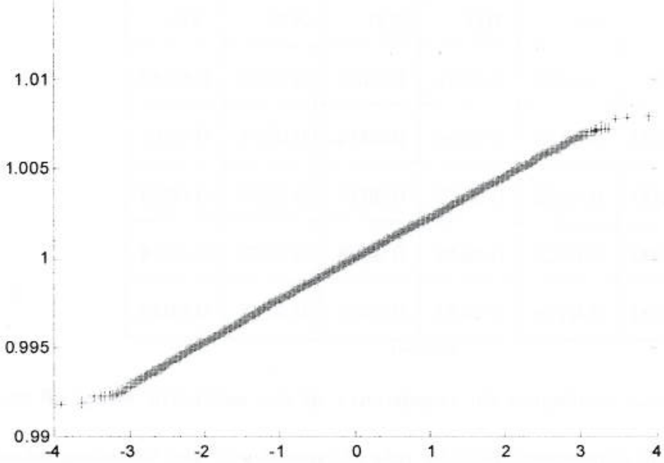

Figure 5: Q-Q plot $\hat{\beta}$ experiment 4

behaviour of its estimator is possible.

The table below reports the results of the Monte Carlo experiment for the standard deviation of $(\hat{\gamma}-\gamma)$ and for the $\mathrm{E}(\hat{\gamma}-\gamma)$:

T

n		10	100	200	300	500
	10	0.0944	0.0313	0.0222	0.0180	0.0139
n	100	0.0153	0.0098	0.0080	0.0057	0.0044
	200	0.0210	0.0071	0.0048	0.0041	0.0031
	300	0.0171	0.0056	0.0040	0.0033	0.0025
	500	0.0130	0.0044	0.0031	0.0029	0.0020

$std(\hat{\gamma}-\gamma)$

T

	10	100	200	300	500
10	0.1393	0.0931	0.0457	0.0321	0.0189
100	0.0153	0.0856	0.0344	0.0311	0.0111
n 200	0.0132	0.0675	0.0311	0.0299	0.0099
300	0.0125	0.0649	0.0298	0.0273	0.0054
500	0.0118	0.0432	0.0276	0.0218	0.0032

$E(\hat{\gamma}-\gamma)$

There are evidences for consistency of the estimator along all the dimensions of the experiment.For the rate of consistency the following estimates are obtained:

$$\begin{aligned} \hat{\varphi} &= 1.1109 \\ \hat{\psi} &= 0.9978 \\ R^2 &= 0.9154 \\ pvalue &= 0.000 \\ Fstatistic &= 63.4478 \\ \hat{\sigma}^2 &= 0.4915 \end{aligned}$$

The Q-Q plot of $\hat{\gamma}$ when (n,T)=(500,500) on top of next page gives evidences for asymptotic normality.

The estimator of $\hat{\gamma}$ is therefore found $\sqrt{nT}$ consistent and asymptotically normal.

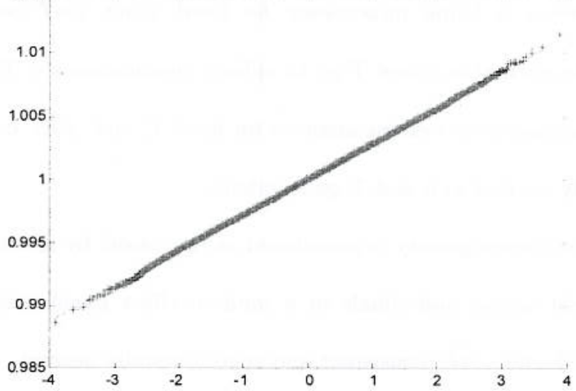

Figure 6: Q-Q plot $\hat{\gamma}$experiment 4

6. Conclusions

This paper presents results from four Monte Carlo experiments on the limiting behaviour of the ols estimator in panels with mixed stationary and unit root regressors.

The investigation has been conducted allowing both the number of cross section observations and the number of time series observations to go to infinity simultaneously.

Evidences from the experiments indicate that,in a general panel model with no endogeneity, the ols estimator of the I(1) regressor is $\sqrt{n}T$ consistent and asymptotically normal,whereas the ols estimator of the stationary regressor is $\sqrt{nT}$ consistent and asymptotically normal.

Once endogeneity is introduced in the model trough a correlation between the regression error and the shock of the unit root process, the estimator of the I(1) regressor is found inconsistent for fixed T,but $\sqrt{nT}$ consistent and asymptotically normal as n and T go to infinity simultaneously .The estimator of the I(0) regressor is as well inconsistent for fixed T, and $\sqrt{nT}$ consistent and asymptotically normal as n and T go to infinity.

If individual heterogeneity is introduced in the model by allowing for a different intercept across individuals in a random effect panel framework both estimator are found $\sqrt{nT}$ consistent and asymptotically normal.

The paper presents no results for individual heterogeneity model with fixed effect. Further investigation on this case would be desirable.

More importantly the paper presents no theoretical results . Monte Carlo experiments cannot provide more than some insight into the behaviour of the estimator. Since no theoretical results are yet available in the literature it is to be hoped that further studies on the properties of the ols estimator in panels with stationary and nonstationary regressors will be undertaken.

References.

References

[1] Baltagi,B(1995),Econometric Analysis of Panel Data.New York:Wiley

[2] Baltagi, B.,Kao, C.,Liu,L.(2008), "Asymptotic Properties of Estimators for the Linear Panel Regression Model with Random Individual Effects and Serially Cor-

related Errors:The Case of Stationary and Non-Stationary Regressors and Residuals."

[3] Baltagi,B,Kramer,W.(1997)," A Simple Linear Trend Model with Error Components",Econometric Theory,13,463-459.

[4] Hsiao,C(1986),Analysis of Panel Data,Cambridge,Cambridge University Press.

[5] Kao,C.(1999), "Spurious Regression and Residual-Based Tests for Cointegrated Regression in Panel Data",Journal of Econometrics,90,1-44

[6] Kao,C.,Emerson,J.(2004),"Testing for Structural Change of a Time Trend Regression In Panel Data",Advances in Econometrics,15,179-222.

[7] Levin,A.,Lin,C.(1993) "Unit Root Tests in Panel Data:New Results",UC San Diego Working Papers.

[8] Pedroni,P(1999),"Panel Cointegration;Asymptotic and Finite Sample Properties of the Pooled Time Series Tests with an Application to the PPP Hypotesis",Indiana University Working Papers in Economics,No.95-013

[9] Pesaran,H.,Smith,R.(1995),"Estimating Long-Run Relationship from Dynamic Heterogeneous Panels",Journal of Econometrics,68,79-113

[10] Phillips,P.C.B.,Moon,H.(1999), "Linear Regression Limit Theory for Nonstationary Panel Data",Econometrica,67,1057-1111.

[11] Phillips,P.C.B,(1986), "Understanding Spurious Regressions in Econometrics",Journal of Econometrics,33,311-340.

[12] Quah,D.(1994)," Exploiting Cross Section Variation for Unit Root Inference in Dynamic Data",Economic Letters,44,9-19.

[13] Robertson,D,Symons,J.(1992)"Some Strange Properties of Panel Data Estimators",Journal of Applied Econometrics,7,175-189

[14] Summers,R,Heston,A.(1991),"The Penn World Table:An Expanded Set of International Comparison 1950-1988",Quarterly Journal of Economics,106,327-368.

[15] Wooldridge,J(2002),Econometric Analysis of Cross Section and Panel Data,Cambridge,MIT Press