Economics
The Open-Access, Open-Assessment E-Journal

Vol. 6, 2012-20 | June 6, 2012 | http://dx.doi.org/10.5018/economics-ejournal.ja.2012-20

# A Note on the Estimation of Long-Run Relationships in Panel Equations with Cross-Section Linkages

*Francesca Di Iorio*
*University of Naples Federico II*
*Stefano Fachin*
*University of Rome "La Sapienza"*

**Abstract** The authors address the issue of estimation and inference in dependent non-stationary panels of small cross-section dimensions. The main conclusion is that the best results are obtained applying bootstrap inference to single-equation estimators, such as fully modified ordinary least squares and dynamic ordinary least squares. Seemingly unrelated regression estimators perform badly, or are even unfeasible, when the time dimension is not very large compared to the cross-section dimension.

**JEL** C15, C23, C33
**Keywords** Panel cointegration; Fully modified ordinary least squares; Fully modified seemingly unrelated regression; Dynamic ordinary least squares; Dynamic seemingly unrelated regression

**Correspondence** Francesca Di Iorio, Dip. TEOMESUS, University of Naples Federico II, via Leopoldo Rodinò 22 - 80138 Naples (Italy); fdiiorio@unina.it; Stefano Fachin, Dip. di Scienze Statistiche, University of Rome "La Sapienza", P.le A. Moro 5, 00185 Roma (Italy); s.fachin@caspur.it

**Citation** Francesca Di Iorio and Stefano Fachin (2012). A Note on the Estimation of Long-Run Relationships in Panel Equations with Cross-Section Linkages. *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 6, 2012-20.
http://dx.doi.org/10.5018/economics-ejournal.ja.2012-20

## 1 Introduction

Consider a panel of $N$ units, with two non stationary variables, (say, $Y$ and $X$) observed over $T$ time periods. In each unit of the panel the two variables are known to be linked by a linear long-run equilibrium (cointegrating) relationship, so that the data generating process (DGP) is the following:

$$y_{it} = \theta_i + \beta_i x_{it} + u^y_{it} \tag{1}$$

$$x_{it} = x_{it-1} + u^x_{it} \tag{2}$$

where $i = 1, \ldots, N, t = 1, \ldots, T$, and $u^x_{it}$ and $u^y_{it}$ are stationary noises. The estimation of (1) when the covariance matrix of the noises is not diagonal, so that the units are dependent, is still a largely unsettled problem. Empirical applications, ignoring efficiency gains, are typically based on single-equation methods (see *e.g.*, Kim et al. (2005), Herzer (2008), Westerlund 2008). This is not surprising, since system estimation with non-stationary variables is fraught with difficulties. Full information maximum likelihood (FIML) methods Groen and Kleibergen (2003) are feasible only when the number of time observations is much larger than the cross-section observations, thus precluding many of the non-stationary panels available in economics and finance. Seemingly unrelated regression (SUR) methods, namely Mark et al. (2005) dynamic SUR (DSUR) and Moon (1999) fully modified SUR (FM-SUR), which are respectively the sytem extensions of dynamic OLS (DOLS) and fully modified ordinary least squares (FM-OLS) are feasible with smaller $T/N$ ratios. However, both require estimation the long-run covariance matrix of the system, a considerably more difficult task (Mark et al. (2005) describe it as ”thorny”) than obtaining the contemporaneuos covariance matrix needed for the baseline SUR. Moon and Perron (2005) claim that SUR estimators are nevertheless superior to single-equation ones in a non-stationary set-up also. However, their simulation study considered a system of very small cross-section size (at most four units with one right-hand side variable, or two units with two variables) and large time dimension ($T = 100$, 300), thus very different from the typical non-stationary panel[1]. This prompts two main questions. First, with empirically relevant sample sizes how large are the efficiency gains (if any) actually delivered by SUR estimators relative to single-equation methods? Should these gains be small, then the widespread use of single-equation estimators would be largely legitimate. Our first goal is thus to compare the estimation performances of single-equation (FM-OLS and DOLS) and SUR system estimators (FM-SUR and DSUR) in panels with small to moderate cross-section dimension and moderate time dimension, characterised by short-run dependence across units. The results will lead to conclusions, hence, advice to practitioners, considerably different from Moon and Perron's.

The second question requires taking a completely different perspective. Efficiency improvements, such as those granted by SUR, are desired in order to have more accurate interval estimation and more reliable tests. Can we reach these targets applying some alternative inference procedure, such has the bootstrap, to standard single-equation estimators? The good simulation results reported for bootstrap inference on FM-OLS (Psaradakis (2001), Fachin (2004)) and unit root and cointegration tests (see *inter alia*, Park (2003), and, for panel extensions, Chang (2004), Fachin (2007), Fuertes 2008) suggest this point is worth investigating.

We shall now first outline the set-up of the Monte Carlo experiment (Section 2) and discuss the results of the comparison between single-equation and SUR estimators (Section 3). In Section 4 we first recall the procedures for bootstrap inference on FM-OLS and then report their performances. Some conclusions are drawn in Section 5.

## 2 Monte Carlo Experiment: Design

The key point here is that the aim of our simulation design cannot be that of obtaining fully general results, as there is a potentially infinite number of dependence structures among the units and variables of a panel. Rather, as mentioned above, we first of all wish to check if the results obtained by Moon and Perron (2005),

[1] For instance, Coakley et al. (2006) describe as typical for macroeconomic panels sample sizes of 20 or 30 cross-section units with from 30 to 100 time observations, corresponding *e.g.*, to about three decades of observations at annual or quarterly frequency for the OECD countries

hold for the sample sizes typical of non-stationary panels. In designing our experiment we will thus follow closely Moon and Perron (2005). The DGP is a simple generalisation of (1)-(2) to the case of $K = 2$ explanatory variables:

$$y_{it} = \theta_i + \beta_{1i}x_{1it} + \beta_{2i}x_{2it} + u^y_{it}, \quad i = 1,\ldots,N, \quad t = 1,\ldots,T, \tag{3}$$

$$x_{kit} = x_{kit-1} + u^x_{kit}, \quad k = 1,2, \quad i = 1,\ldots,N, \quad t = 1,\ldots,T. \tag{4}$$

where $u^x_{kit}, u^y_{it}$ are $I(0)$ noises, so that both the $x's$ and $y$ are $I(1)$. In the non-stationary panels literature it is quite common (see, *e.g.*, Pesaran 2006) to introduce some realism in the simulation design through parameters heterogenous across units. Here we will follow this practice, generating the regression coefficients respectively as $\theta_i \sim U(2,4)$ and as $\beta_{ki} \sim U(1,3)$, where $k = 1,2$. The same set of coefficients has been used for all Monte Carlo replications. It should be remarked that, provided the error variances are suitably controlled to keep the signal-noise ratio constant, the use of heterogenous parameters instead of the homogenous ones used by Moon and Perron (2005) has no consequences on the performances of estimators which allow for heterogeneity[2]. Things are obviously different for pooled estimators, which are misspecified under heterogeneity. Since this class of estimators will not be examined in our experiment the point is irrelevant.
The errors of equations (3) and (4) are drawn from a Multivariate Normal distribution with non-diagonal covariance matrix, so that there is short-run dependence across equations and units (the case of long-run dependence is ruled out, as FM-SUR, which require the inversion of the long-run covariance matrix would then not be feasible). More precisely, letting $\mathbf{u}^x_t = [\mathbf{u}^{x\prime}_{1t}\mathbf{u}^{x\prime}_{2t}\ldots\mathbf{u}^{x\prime}_{Nt}]'$, where $\mathbf{u}^{x\prime}_{it} = [u^x_{1it}u^x_{2it}]'$ and $\mathbf{u}^y_t = [u^y_{1t}u^y_{2t}\ldots u^y_{Nt}]'$, we have

$$\begin{bmatrix} \mathbf{u}^y_t \\ \mathbf{u}^x_t \end{bmatrix}_{(N+2N)\times 1} \sim \quad iidN\left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} R & \Delta \\ \Delta' & \Phi \end{bmatrix}_{(N+2N)\times(N+2N)} \right), \tag{5}$$

where $R$ is a full $N \times N$ matrix governing the dependence across units in the $u^{y\prime}_{it}$s, $\Delta$ is a $N \times 2N$ matrix governing the dependence between the $u^x$ and $u^y$ noises, and

[2] The results of the simulazion with homogenous parameters, not included here for sake of brevity, is obviously available on request.

finally $\Phi$ is a $2N \times 2N$ matrix governing the dependence in the $u^{x\prime}s$ within and across units.

Since Moon and Perron report the performances of both FM-OLS and FM-SUR estimators to be negatively affected by the degree of endogeneity of the $X$'s we decided to control this dimension of the experiment accurately, imposing an homogeneous endogeneity parameter $\delta$ and running two sets of experiments with $\delta = 0.2$ and $\delta = 0.4$. In both cases the $\Delta$ matrix has a block form ensuring that there is constant correlation between the noise of any $X$ and that of the relevant $Y$ equation, and no correlation across units:

$$\Delta_{N\times 2N} = \begin{bmatrix} \delta & \delta & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \delta & \delta & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \dots & \delta & \delta \end{bmatrix}. \tag{6}$$

We instead allow some heterogeneity across units in the dependence parameters, the elements of the $\Phi$ matrix. Without loss of generality, we assume $x_{1it}$ and $x_{2it}$ to be incorrelated. Letting $\phi_{lk}^{(ij)} = cov(u_{li}^x, u_{kj}^x)$ denote the covariance between the noise of the variable $X_l$ in the $i^{th}$ unit and that of the variable $X_k$ in the $j^{th}$ unit, we have:

$$\Phi_{2N\times 2N} = \begin{bmatrix} 1 & 0 & \phi_{11}^{(12)} & \phi_{12}^{(12)} & \dots & \phi_{11}^{(1N)} & \phi_{12}^{(1N)} \\ 0 & 1 & \phi_{21}^{(12)} & \phi_{22}^{(12)} & \dots & \phi_{21}^{(1N)} & \phi_{22}^{(1N)} \\ \phi_{11}^{(21)} & \phi_{12}^{(21)} & 1 & 0 & \dots & \phi_{11}^{(2N)} & \phi_{12}^{(2N)} \\ \phi_{21}^{(21)} & \phi_{22}^{(21)} & 0 & 1 & \dots & \phi_{21}^{(2N)} & \phi_{22}^{(2N)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi_{11}^{(N1)} & \phi_{12}^{(N1)} & \phi_{11}^{(N2)} & \phi_{12}^{(N2)} & \dots & 1 & 0 \\ \phi_{21}^{(N1)} & \phi_{22}^{(N1)} & \phi_{21}^{(N2)} & \phi_{22}^{(N2)} & \dots & 0 & 1 \end{bmatrix} \tag{7}$$

with $\phi_{lk}^{(ij)} \sim U(0.3, 0.4)$. The off-diagonal elements of $R$, $\rho_{ij}$, are also generated as $U(0.3, 0.4)$, while $\rho_{ii} = 1\ \forall i$. Again, the parameters thus generated have been kept fixed across the Monte Carlo repetitions.

The time and cross-section sample sizes have been chosen trying to strike a balance between empirical relevance (as mentioned above, most macroeconomic panels have $N$ around 20 or 30 and $T$ often much smaller than 100) and the requirements of the SUR estimator, which is feasible only with rather large $T/N$ ratios. We thus fixed $N = 5, 10$ and $T = 50, 100$. Finally, we set the number of Monte Carlo simulations ($M$) to 1000.

## 3 Simulation Results: Comparison of Single-Equation and SUR Estimators

In Tables 1 and 2 we report some summary statistics describing the results obtained estimating (3) by single equation (FM-OLS and DOLS) and system (FM-SUR and DSUR) methods. To define the expressions for these estimators we need some notation. Following Moon (1999), for the simple bivariate case (1)-(2) first of all define $\omega_t = (u_t^y, u_t^x)'$ and assume that $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \omega_t \to B(r)$ . The long-run covariance matrix of $B(r)$ is $\Omega = \sum_{h=-\infty}^{\infty} E(\omega_0 \omega_h')$, and the one-sided long-run covariance matrix $\Xi = \sum_{h=0}^{\infty} E(\omega_0 \omega_h')$, both partitioned in the usual way as

$$\Omega = \begin{pmatrix} \Omega_{yy} & \Omega_{yx}' \\ \Omega_{yx} & \Omega_{xx} \end{pmatrix}, \quad \Xi = \begin{pmatrix} \Xi_{yy} & \Xi_{yx}' \\ \Xi_{yx} & \Xi_{xx} \end{pmatrix}$$

where all blocks have dimension $N \times N$. Further, let $X_t = diag(x_{1t}, \ldots, x_{Nt}), Y_t = (y_{1t}, \ldots, y_{Nt})'$, $\Omega_{yy.x} = (\Omega_{yy} - \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xy})$ . Denoting by an hat a consistent estimate, then:

$$\hat{y}_t^+ = y_t - \hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1}\Delta x_t \ , \quad \widetilde{Y}_t^+ = (\tilde{y}_{1t}^+, \ldots, \tilde{y}_{Nt}^+)'$$
$$\tilde{y}_{it}^+ = y_{it} - \hat{\Omega}_{yx}^{ii}(\hat{\Omega}_{xx}^{ii})^{-1}\Delta x_{it} \ , \quad i = 1, \ldots, N.$$

Also,

$$\widehat{\xi} = (\widehat{\xi}_1', \ldots, \widehat{\xi}_n')' \ , \quad \widehat{\xi}_i = \widehat{\Xi}_{xy}^{ii} - \hat{\Omega}_{yx}^{ii}(\hat{\Omega}_{xx}^{ii})^{-1}\widehat{\Xi}_{xx}^{ii}$$

$$\hat{\pi}_i = \widehat{\Xi}_{xx}^{i.}((\hat{\Omega}_{yx}^{ii}\hat{\Omega}_{xx}^{ii})_{i.}^{-1})' \ ,$$

and, finally,

$$\widehat{\psi}_i = (\widehat{\Xi}_{xy}^{i.}(\hat{\Omega}_{yy.x}^{-1})_{i.})' - (\widehat{\xi}_{xx}^{i.}(\hat{\Omega}_{yy.x}^{-1}\hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1})_{i.})'$$

with $\widehat{\psi} = (\widehat{\psi}_1', \ldots, \widehat{\psi}_N')'$ . The estimators are then defined as follows:

$$\hat{\beta}^{FM-OLS} = \left( \sum_{t=1}^{T} X_t X_t{}' \right)^{-1} \left( \sum_{t=1}^{T} X_t \widetilde{Y}_t^{+} - T \widehat{\xi}_i \right)$$

$$\hat{\beta}^{DOLS} = \left( \sum_{t=p+1}^{T-p} X_t X_t' \right)^{-1} \left( \sum_{t=p+1}^{T-p} X_t \widetilde{Y}_t^{+} \right)$$

$$\hat{\beta}^{FM-SUR} = \left( \sum_{t=1}^{T} X_t \hat{\Omega}_{yy.x}^{-1} X_t' \right)^{-1} \left( \sum_{t=1}^{T} X_t \hat{\Omega}_{yy.x}^{-1} \widehat{Y}_t^{+} - T \widehat{\psi} \right)$$

$$\hat{\beta}^{DSUR} = \left( \sum_{t=p+1}^{T-p} X_t^{-1} \hat{\Omega}_{yy}^{-1} X_t' \right)^{-1} \left( \sum_{t=p+1}^{T-p} X_t \hat{\Omega}_{yy}^{-1} Y_t \right)$$

where $\widehat{\beta} = (\widehat{\beta}_1', \ldots, \widehat{\beta}_N')$, and $\widehat{\beta}_i = (\theta_i, \beta_i)$. Small sample point estimation performance of the estimators is usually evaluted by simulation on the basis of the mean over the $M$ simulations of the relative bias, $M^{-1} \sum_{m=1}^{M} (\widehat{\beta}_m - \beta) \beta^{-1}$ . In a DGP such as (3)-(4), with $N$ units and $K$ explanatory variables, we evaluate overall point estimation performance by the average over units and parameters of the absolute value (so to avoid compensating errors in opposite directions) of the bias of the estimates of each parameter:

$$\overline{bias} = (KN)^{-1} \sum_{k=1}^{K} \sum_{i=1}^{N} |M^{-1} \sum_{m=1}^{M} (\widehat{\beta}_{kim} - \beta_{ki}) \beta_{ki}^{-1}| \tag{8}$$

Dispersion is analogously measured by the mean over the $N$ units and $K$ parameters of the relative Monte Carlo standard error, $\left( \sqrt{M^{-1} \sum_{m=1}^{M} (\widehat{\beta}_{kim} - \overline{\widehat{\beta}}_{ki})^2} \right) \beta_{ki}^{-1}$:

$$\overline{s.e.} = (KN)^{-1} \sum_{k=1}^{K} \sum_{i=1}^{N} \left[ \left( \sqrt{M^{-1} \sum_{m=1}^{M} (\widehat{\beta}_{kim} - \overline{\widehat{\beta}}_{ki})^2} \right) \beta_{ki}^{-1} \right] \tag{9}$$

In our experiment we will also evaluate testing performances. Given the extremely different size performances of the single-equation and system estimators, we

will concentrate on Type I errors. We thus tested the hypothesis $H_0 : \beta_{ki} = \beta_{ki}^{(0)}$, $k = 1,\ldots,K$, $i = 1,\ldots,N$, where $\beta_{ki}^{(0)}$ is the value of the slope parameter used in the Monte Carlo DGP.
The first remark in order is that the all estimators are indeed more biased in DGP's with an higher degree of endogeneity.
The second, very important, remark is that the SUR procedure turned out to be practically unfeasible for $T = 50$ and $N = 10$. The covariance matrix, although not exactly singular, was always so ill-conditioned that the estimators turned out highly numerically unstable even using a generalised Moore–Penrose inversion routine[3]. Hence, we do not report results for this $(T,N)$ combination. Since these samples sizes are rather common in applied work on non-stationary panels (with indeed the time sample often actually smaller than this one) this is an important finding.

Let us now go into some detail, considering point estimation first. All estimators are essentially unbiased even with the smaller time sample. However, from the first two columns of Table 1 we can appreciate that FM-OLS delivers a slightly better performance than DOLS, while the ranking of the two SUR estimators is not obvious (considering also that DSUR, contrary to FM-SUR, could be computed also for the $T = 50$, $N = 10$ combination). SUR estimators tend to be somehow more biased than the OLS ones. For instance, when $T = 50$, $N = 5$ and $\delta = 0.2$ (first raw of Tables 1 and 2) the average relative bias is essentially the same for FM-OLS, DOLS and FM-SUR (respectively, 0.32%, 0.32% and 0.33%) and higher for DSUR (0.45%); when $T = 100$, $N = 10$ and $\delta = 0.4$ (last raw of Tables 1 and 2) the bias of FM-SUR is larger than that of FM-OLS (0.51% against 0.16%), and DSUR (0.41%) which is even less biased than its single-equation counterpart (0.46%).

[3] It should be remarked that the problem here, estimating the long-run covariance matrix, is considerably more difficult than the standard SUR problem examined by Foschi et al. (2003)

*Table* 1: Single-Equation Estimation
Performance of Fully Modified and Dynamic OLS Estimators

| | | | $\overline{bias}$ | | $\overline{s.e.}$ | | $\widehat{\alpha}$ | |
|---|---|---|---|---|---|---|---|---|
| $\delta$ | $T$ | $N$ | $FM$ | $D$ | $FM$ | $D$ | $FM$ | $D$ |
| 0.2 | 50 | 5 | 0.32 | 0.32 | 3.37 | 3.59 | 13.23 | 12.71 |
| | | 10 | 0.24 | 0.59 | 4.54 | 4.62 | 14.19 | 19.61 |
| | 100 | 5 | 0.10 | 0.15 | 1.87 | 1.94 | 9.18 | 11.31 |
| | | 10 | 0.10 | 0.23 | 1.89 | 1.98 | 9.51 | 14.52 |
| 0.4 | 50 | 5 | 0.65 | 0.67 | 3.36 | 3.64 | 13.42 | 13.33 |
| | | 10 | 0.47 | 1.21 | 4.48 | 4.35 | 14.11 | 19.88 |
| | 100 | 5 | 0.19 | 0.32 | 1.86 | 1.96 | 9.49 | 11.65 |
| | | 10 | 0.16 | 0.46 | 1.87 | 1.98 | 9.42 | 15.25 |

$FM$: Fully Modified OLS; $D$: Dynamic OLS;

$\overline{bias}$: see (8); values $\times 100$;
$\overline{s.e.}$: see (9); values $\times 100$;
$\widehat{\alpha}$: rejection rates ($\times 100$) of tests of nominal size 5%, $H_0$ true;

The Monte Carlo variance is approximately the same for FM and dynamic estimators, with both SUR estimators always less precise than the OLS ones. As to be expected, the variance falls with $T$ and increases with $N$ for given time sample.

Overall, FM-OLS dominates DOLS in terms of both bias and dispersion, while FM-SUR and DSUR seems to be largely equivalent in the cases when the former could be computed. System estimators are somehow more biased and less efficient than single-equation ones, but the differences are small (and not always in this direction). However, if we turn to hypothesis testing (last two columns of Table 1 and 2) we discover that the performance of FM-OLS and DOLS, although disappointing, is vastly superior to that of both FM-SUR and DSUR, with the former simply disastrous. While the Type I errors of FM-OLS fall between 9% and 14%, those of FM-SUR are about twice as large, falling between 19% (the two cases with $T = 100$, $N = 5$) and over 40% (all the other $T,N$ combinations). The

DSUR estimator seems to be more robust, with Type I errors sometimes close to those of DOLS and generally much smaller than those of FM-SUR (for instance, when $T = 100$, $N = 10$, $\delta = 0.4$, last raw of Table 2, 12.96% for DSUR and 43.87% for FM-SUR).

*Table* 2: System Estimation
Performance of Fully Modified and Dynamic SUR Estimators

| | | | $\overline{bias}$ | | $\overline{s.e.}$ | | $\widehat{\alpha}$ | |
|---|---|---|---|---|---|---|---|---|
| $\delta$ | $T$ | $N$ | $FM$ | $D$ | $FM$ | $D$ | $FM$ | $D$ |
| 0.2 | 50 | 5 | 0.33 | 0.45 | 4.03 | 4.54 | 45.78 | 16.35 |
| | | 10 | - | 0.56 | - | 5.87 | - | 23.93 |
| | 100 | 5 | 0.14 | 0.16 | 1.94 | 2.16 | 18.69 | 10.57 |
| | | 10 | 0.24 | 0.20 | 2.15 | 2.21 | 41.67 | 12.38 |
| 0.4 | 50 | 5 | 0.49 | 0.57 | 4.08 | 4.46 | 47.46 | 54.95 |
| | | 10 | - | 1.14 | - | 5.90 | - | 24.63 |
| | 100 | 5 | 0.31 | 0.32 | 1.95 | 2.16 | 19.80 | 23.30 |
| | | 10 | 0.51 | 0.41 | 2.17 | 2.22 | 43.87 | 12.96 |

-: not available (numerical overflow); all values $\times 100$;
*symbols and abbreviations*: see Table 1

The reason for the extremely poor performance of FM-SUR is not obvious from the average bias and Monte Carlo variability reported in Table 2. To shed some light on the problem we need to examine in detail the estimation performances for each unit. In Table 3 we report bias and variability statistics for each unit of the combination $T = 50$, $N = 5$, $\delta = 0.4$; those for other parameter combinations $(N = 10, T = 100, \delta = 0.4)$ are similar, and thus not reported (as customary, they are available on request). The most important finding is that the standard asymptotic formulas for the variance of the FM-SUR estimator always grossly underestimate its actual variability: for instance, the Monte Carlo standard error ($\times 100$, but not normalised on the coefficient value, differently from (9) used above) of the estimates of $\beta_1$ is about 0.13 in units 3, 4 and 5, but the asymptotic formulas yield in all three cases estimates less than 0.05; the average across units and coefficients of the Monte Carlo standard errors is about 0.08, that of the asymptotic ones less than 0.03. Since asymptotic inference is based upon these grossly underestimated variances its poor performance is not surprising. On the other hand, the asymptotic estimates of the standard errors of the DSUR estimator (on average, 0.065) are much closer to the Monte Carlo standard errors (on average, 0.088). This explains why the performance of asymptotic hypothesis testing on DSUR, though generally rather poor, is relatively better than that on FM-SUR.

*Table* 3: FM-SUR and DSUR Estimators - Bias and Variability for Individual Units

| | | $FM-SUR$ | | | $DSUR$ | | |
|---|---|---|---|---|---|---|---|
| *Unit* | *parameter* | *bias* | *MC s.e.* | $\overline{\widehat{\sigma}}$ | *bias* | *MC s.e.* | $\overline{\widehat{\sigma}}$ |
| 1 | $\beta_1$ | 0.010 | 0.052 | 0.022 | 0.011 | 0.057 | 0.044 |
| | $\beta_2$ | 0.013 | 0.081 | 0.030 | 0.011 | 0.072 | 0.053 |
| 2 | $\beta_1$ | 0.004 | 0.032 | 0.014 | 0.010 | 0.050 | 0.038 |
| | $\beta_2$ | 0.014 | 0.068 | 0.027 | 0.016 | 0.067 | 0.050 |
| 3 | $\beta_1$ | 0.003 | 0.134 | 0.043 | 0.013 | 0.114 | 0.084 |
| | $\beta_2$ | 0.009 | 0.065 | 0.023 | 0.003 | 0.061 | 0.045 |
| 4 | $\beta_1$ | 0.046 | 0.132 | 0.045 | 0.022 | 0.111 | 0.083 |
| | $\beta_2$ | 0.003 | 0.046 | 0.017 | 0.009 | 0.059 | 0.044 |
| 5 | $\beta_1$ | 0.021 | 0.140 | 0.046 | 0.078 | 0.231 | 0.167 |
| | $\beta_2$ | 0.019 | 0.052 | 0.021 | 0.004 | 0.055 | 0.043 |
| | *mean* | 0.014 | 0.080 | 0.029 | 0.018 | 0.088 | 0.065 |

*DGP*: $T = 50, \delta = 0.4$;
$bias = M^{-1}\sum_m^M(\widehat{\beta}_{kim} - \beta_{ki}), k = 1,2, i = 1,\ldots,5.$
$MC\ s.e. = 100 \times \sqrt{M^{-1}\sum_m^M(\widehat{\beta}_{kim} - \overline{\widehat{\beta}}_{ki})^2}$;
$\overline{\widehat{\sigma}}$ : average estimated asymptotic standard error $\times 100$;
*other symbols and details*: see Table 1.

Summing up, with small to moderate $T/N$ ratios system estimators (*i*) may be even unfeasible, (*ii*) when feasible, are generally more biased and less efficient than single-equation estimators, and, finally, (*iii*) are associated with strongly oversized asymptotic tests.

The best option seems to be FM-OLS, which deliver the best results in terms of bias, efficiency and Type I errors of asymptotic tests. The latter, however, are significantly higher than nominal significance levels, which implies that Gaussian confidence intervals will be deceivingly short and have coverage smaller than nominal. Can we do any better? To this end, in the next section we will recall some standard bootstrap procedures for inference on FM-OLS and present the results of a small simulation experiment.

## 4 Bootstrap Inference in Cointegrating Regressions

Although of rather recent introduction, bootstrap inference in regressions with $I(1)$ variables is now well established (see *e.g.*, Herwartz and Neumann (2005), Chang and Song 2006). Since the details are beyond the scope of this paper we will now simply sketch the basic concepts in order to establish notation.

For ease of exposition, consider the simple case of equations (1)-(2). Bootstrap inference involves two key steps: first, constructing the pseudo-data sets; second, defining the test statistics or confidence intervals to be used. Let us examine them in turn.

Denoting by $u_{it}^{y*}$ the bootstrap noise, which we will discuss below, in the case of hypothesis testing the systematic part of the bootstrap DGP is given by the null hypothesis to be tested (say, $H_{0:}\ \beta_i = \beta_i^{(0)}$):

$$y_{it}^* = \widehat{\theta}_i + \beta_i^{(0)} x_{it} + u_{it}^{y*} \tag{10}$$

while for interval estimation we use the unconstrained parameter estimates:

$$y_{it}^* = \widehat{\theta}_i + \widehat{\beta}_i x_{it} + u_{it}^{y*} \tag{11}$$

The key point to take into account when generating the bootstrap noise $u_{it}^{y*}$ is the presence of dependence both in the time series and in the cross-section dimensions.

The former aspect has been the subject of the vast debate, whose details are again beyond the scope of this paper; for a review, see Politis (2003). Essentially, we can either follow a model-based (parametric) approach or a non-parametric one. In the former case in a first step the data are filtered through AR models, so to obtain white-noise residuals to be resampled. In the latter approach blocks of observations of length proportional to the memory of the series and random starting point are drawn with replacement from the dependent series. Here, as in Di Iorio and Fachin (2007) we will follow the latter approach: a block bootstrap algorithm, the Stationary Bootstrap by Politis and Romano (1994), will be applied to the unconstrained residuals delivered by FM-OLS estimation of (3). In principle a critical point of block bootstrap methods is the choice of block length. In our simulations we kept the block length always fixed at $T/10$, a value which give good results in the simulations by Paparoditis and Politis (2003), where the issue is widely discussed. In either case, to preserve the cross-unit dependence structure we simply need to resample the entire $T \times N$ matrix of residuals. In other terms, the resampling algorithm swaps (blocks of) rows but keeps the columns fixed in their positions, so that in the bootstrap data set the dependence structure across the columns of the original data set is reproduce exactly as it is.

As usual, the bootstrap estimate of the $p$-value of two-tailed tests will be $p_i^* = prop(|t_{ib}^*| > t_i)$, with $t_i$ the usual $t$-statistic $t_i = s_{\beta_i}^{-1}(\widehat{\beta}_i - \beta_i^{(0)})$, $t_{ib}^* = s_{\beta_i}^{*-1}(\widehat{\beta}_{ib}^* - \widehat{\beta}_i)$, $\widehat{\beta}_{ib}^*$ the FM-OLS estimate of $\beta_i$ computed on the $b^{th}$ pseudo-data set ($b = 1, \ldots, B$), $s_{\beta_i}$ and $s_{\beta_i}^*$ the estimated standard errors of the estimators of the actual and bootstrap samples, respectively. One simple way to compute confidence intervals is to take the desired quantiles of the distribution of the $\widehat{\beta}_{ib}^{*\prime}$s. An $\alpha$-level confidence interval for $\beta_i$ may then simply be given by

$$[Q_{\alpha/2}(\widehat{\beta}_i^*), Q_{1-\alpha/2}(\widehat{\beta}_i^*)] \tag{12}$$

where $\widehat{\beta}_i^* = \left[\widehat{\beta}_{i1}^* \ldots \widehat{\beta}_{iB}^*\right]$. In principle basing the interval on a pivotal quantity should deliver better results. Psaradakis (2001) thus suggests the percentile-$t$ interval

$$[\widehat{\beta}_i - Q_{1-\alpha/2}(\mathbf{t}_b^*) s_{\beta_i}, \widehat{\beta}_i - Q_{\alpha/2}(\mathbf{t}_b^*) s_{\beta_i}] \tag{13}$$

where the Gaussian quantiles used in asymptotic inference are replaced by those of the bootstrap distribution (empirical estimate of the unknown small sample distribution of the studentized statistic). The superiority of the second type of interval depends entirely upon the quality of the estimates of the standard errors (see *e.g.*, Kilian 1999). Hence, in our study we shall compute both type of intervals.

Let us now turn to the simulation results. We saw above that in our set-up Gaussian inference on FM-OLS results in confidence intervals with large size bias and coverage smaller than nominal; from Table 4 we can appreciate that both problems are partially solved applying rather standard bootstrap methods. Although both bootstrap confidence intervals have coverage smaller than nominal (between 89% and 92% for the simple interval and between 90% and 94% for the studentised interval, instead of the desired 95%) and the test is undersized (Type I errors between 1% and 4% for a 5% test), both indicators are measures are markedly better than those of Gaussian inference (Type I errors between 9% and 14%, and, as a consequence, confidence intervals with coverage between 86% and 91%; see Table 1). The question if we can further improve on these performances (for instance, by fine-tuning the block size; see Politis and White 2004) is beyond the scope of this paper. Here the point is simply that applying well-established bootstrap methods to FM-OLS we can improve over standard Gaussian inference in terms of coverage and Type I errors, whereas moving to SUR system estimation all results worsen.

*Table* 4: Bootstrap Inference on FM-OLS

| | $\delta = 0.2$ | | | | $\delta = 0.4$ | | | |
|---|---|---|---|---|---|---|---|---|
| $T$ | 50 | | 100 | | 50 | | 100 | |
| $N$ | 5 | 10 | 5 | 10 | 5 | 10 | 5 | 10 |
| *simple* | 89.7 | 89.5 | 89.7 | 88.6 | 90.1 | 89.8 | 90.4 | 89.3 |
| $t$ | 90.7 | 90.0 | 90.7 | 92.0 | 90.8 | 89.9 | 91.2 | 92.1 |
| $\widehat{\alpha}$ | 7.9 | 8.1 | 7.9 | 9.3 | 8.1 | 8.0 | 8.5 | 9.0 |

*simple*: simple percentile 5% confidence interval, see (12);
$t$: studentised 5% confidence interval, see (13);
$\widehat{\alpha}$: Type I error, nominal significance level 5%;
*Bootstrap settings*: 1000 redrawings; block size $T/10$.

## 5 Conclusions

The Monte Carlo analysis conducted in this paper compares single equation (FM-OLS and DOLS) and system (FM-SUR and DSUR) estimators of long-run relationships in panels under more realistic time series and cross-section dimensions than previous studies. The Monte Carlo results unambiguously suggest that single-equation FM-OLS alongside a block-bootstrap method provides more accurate estimation and inference. These conclusions, in stark contrast to Moon and Perron (2005), should not come as a surprise. As remarked by Mark et al. (2005), the properties of SUR estimators depend critically upon the quality of the estimate of the covariance matrix. This task may be easy in panels with a very small cross-section relative to the time-series dimension, such as those examined by Moon and Perron (2005), but is typically difficult in even slightly larger cross-section panels, such as those considered in our study.

**Acknowledgement** Financial support from the Department of Statistics of the University of Naples Federico II, University of Rome "La Sapienza" and MIUR is gratefully acknowledged. Special thanks to Christoph Hanck for posting a comment pointing out a mistake in the discussion paper. The usual disclaimers apply.

## References

Chang, Y. (2004). Bootstrap unit root tests in panels with cross-sectional dependency. *Journal of Econometrics*, 120(2): 263–293. DOI 10.1016/S0304-4076(03)00214-8. URL http://dx.doi.org/10.1016/S0304-4076(03)00214-8.

Chang, Y. a. J. P., and Song, K. (2006). Bootstrapping cointegrating regressions. *Journal of Econometrics*, 133(2): 703–739. DOI 10.1016/j.jeconom.2005.06.011. URL http://dx.doi.org/10.1016/j.jeconom.2005.06.011.

Coakley, J., Fuertes, A.-M., and Smith, R. (2006). Unobserved heterogeneity in panel time series models. *Computational Statistics & Data Analysis*, 50(9):

2361–2380. DOI 10.1016/j.csda.2004.12.015. URL http://dx.doi.org/10.1016/j.csda.2004.12.015.

Di Iorio, F., and Fachin, S. (2007). Testing for breaks in cointegrated panels - with an application to the Feldstein-Horioka Puzzle. *Economics: The Open-Access, Open-Assessment E-Journal*, 1(2007-14). DOI 10.5018/economics-ejournal.ja.2007-14. URL http://dx.doi.org/10.5018/economics-ejournal.ja.2007-14.

Fachin, S. (2004). Bootstrap inference on fully modified estimates of cointegrating coefficients: A comment. *Economics Bulletin*, 3(3): 1–8. URL http://www.accessecon.com/pubs/EB/2004/Volume3/EB-04C20015A.pdf.

Fachin, S. (2007). Long-run trends in internal migrations in italy: A study in panel cointegration with dependent units. *Journal of Applied Econometrics*, 22(2): 401–428. DOI 10.1002/jae.907. URL http://dx.doi.org/10.1002/jae.907.

Foschi, P., Belsley, D., and Kontoghiorghes, E. (2003). A comparative study of algorithms for solving seemingly unrelated regressions models. *Computational Statistics & Data Analysis*, 44(12): 3–35. DOI 10.1016/S0167-9473(03)00028-8. URL http://www.sciencedirect.com/science/article/pii/S0167947303000288.

Fuertes, A.-M. (2008). Sieve bootstrap -tests on long-run average parameters. *Computational Statistics & Data Analysis*, 52(7): 3354–3370. DOI 10.1016/j.csda.2007.11.014. URL http://www.sciencedirect.com/science/article/pii/S0167947307004495.

Groen, J., and Kleibergen, F. (2003). Likelihood-based cointegration analysis in panels of vector error-correction models. *Journal of Business & Economic Statistics*, 21(2): 295–318. URL http://EconPapers.repec.org/RePEc:bes:jnlbes:v:21:y:2003:i:2:p:295-318.

Herwartz, H., and Neumann, M. (2005). Bootstrap inference in systems of single equation error correction models. *Journal of Econometrics*, 128(1): 165–193. DOI 10.1016/j.jeconom.2004.08.011. URL http://www.sciencedirect.com/science/article/pii/S0304407604001538.

Herzer, D. (2008). The long-run relationship between outward FDI and domestic output: Evidence from panel data. *Economics Letters*, 100(1): 146–149. DOI 10.1016/j.econlet.2007.12.004. URL http://www.sciencedirect.com/science/article/pii/S0165176507004302.

Kilian, L. (1999). Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses. *Review of Economics and Statistics*, 81(4): 652–660. ISSN 0034-6535. DOI 10.1162/003465399558517. URL http://dx.doi.org/10.1162/003465399558517.

Kim, H., Oh, K.-Y., and Jeong, C.-W. (2005). Panel cointegration results on international capital mobility in Asian economies. *Journal of International Money and Finance*, 24(1): 71–82. DOI 10.1016/j.jimonfin.2004.10.006. URL http://www.sciencedirect.com/science/article/pii/S0261560604000932.

Mark, N., Ogaki, M., and Sul, D. (2005). Dynamic seemingly unrelated cointegrating regressions. *The Review of Economic Studies*, 72(3): 797–820. DOI 10.1111/j.1467-937X.2005.00352.x. URL http://restud.oxfordjournals.org/content/72/3/797.abstract.

Moon, H. (1999). A note on fully-modified estimation of seemingly unrelated regressions models with integrated regressors. *Economics Letters*, 65(1): 25–31. DOI 10.1016/S0165-1765(99)00129-9. URL http://www.sciencedirect.com/science/article/pii/S0165176599001299.

Moon, H., and Perron, B. (2005). Efficient estimation of the seemingly unrelated regression cointegration model and testing for purchasing power parity. *Econometric Reviews*, 23(4): 293–323. DOI 10.1081/ETC-200040777. URL http://www.tandfonline.com/doi/abs/10.1081/ETC-200040777.

Paparoditis, E., and Politis, D. (2003). Residual-based block bootstrap for unit root testing. *Econometrica*, 71(3): 813–855. DOI 10.1111/1468-0262.00427. URL http://dx.doi.org/10.1111/1468-0262.00427.

Park, J. (2003). Bootstrap unit root tests. *Econometrica*, 71(6): 1845–1895. ISSN 1468-0262. DOI 10.1111/1468-0262.00471. URL http://dx.doi.org/10.1111/1468-0262.00471.

Pesaran, M. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4): 967–1012. URL http://www.jstor.org/stable/3805914.

Politis, D. (2003). The impact of bootstrap methods on time series analysis. *Statistical Science*, 18(2): 219–230. URL http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?handle=euclid.ss/1063994977\&view=body\&content-type=pdf_1.

Politis, D., and Romano, J. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428): 1303–1313. URL http://www.jstor.org/stable/2290993.

Politis, D., and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1): 53–70. DOI 10.1081/ETC-120028836. URL http://www.tandfonline.com/doi/abs/10.1081/ETC-120028836.

Psaradakis, Z. (2001). On bootstrap inference in cointegrating regressions. *Economics Letters*, 72(1): 1–10. DOI 10.1016/S0165-1765(01)00410-4. URL http://www.sciencedirect.com/science/article/pii/S0165176501004104.

Westerlund, J. (2008). Panel cointegration tests of the Fisher effect. *Journal of Applied Econometrics*, 23(2): 193–233. DOI 10.1002/jae.967. URL http://dx.doi.org/10.1002/jae.967.