Working Paper Series
Department of Economics
University of Verona

Moment Conditions and Neglected Endogeneity in Panel Data Models

Giorgio Calzolari, Laura Magazzini

WP Number: 2 February 2011

ISSN: 2036-2919 (paper), 2036-4679 (online)

Moment Conditions and Neglected Endogeneity in Panel Data Models

Giorgio Calzolari* Laura Magazzini†‡

February 7, 2011

Abstract

This paper develops a new moment condition for estimation of linear panel data models. When added to the set of instruments devised by Anderson, Hsiao (1981, 1982) for the dynamic model, the proposed approach can outperform the GMM methods customarily employed for estimation. The proposal builds on the properties of the iterated GLS, that, contrary to conventional wisdom, can lead to a consistent estimator in particular cases where endogeneity of the explanatory variables is neglected. The targets achieved are a reduction in the number of moment conditions and a better performance over the most widely adopted techniques.
Keywords: panel data, dynamic model, GMM estimation, endogeneity.

1 The linear panel data model

Consider the linear panel data model on N units observed over $T \geq 2$ time periods:

$$y_{it} = x'_{it}\beta + \varepsilon_{it}, \quad i = 1, ..., N; t = 1, ..., T \tag{1}$$

where x_{it} can also contain lagged values of the y_{it} or (expectation of) leading values. Our analysis focuses on micro panels where N is typically large and T is typically small.

*Department of Statistics, Università di Firenze, <calzolar@ds.unifi.it>

†Department of Economics, Università di Verona, <laura.magazzini@univr.it>

‡We gratefully acknowledge comments and suggestions from E. Battistin, G. Carmeci, R.J. Lucchetti, F. Peracchi, A. Sembenelli, and conference participants at the Fourth Italian Congress of Econometrics and Empirical Economics, Pisa, January 19-21, 2011. Of course, we retain full responsibility for the contents of this paper.

The error term ε_{it} is usually decomposed into two sources: $\alpha_i + e_{it}$, where α_i captures individual 'unobserved heterogeneity', constant over time and different across units, and e_{it} is an idiosyncratic component changing both over time and across units.[1]

Since the early Sixties two approaches have been considered for estimation in a static framework, i.e. the *fixed effect* (FE) and the *random effect* (RE) estimation (Mundlak, 1978).[2] The FE approach to estimation removes α_i from the estimating equation by considering the within-group transformation. Within the RE framework, α_i is included in the error term, whose variance is estimated. In the latter case, a generalized least squares (GLS) estimator is usually adopted for β. The difference between the two approaches is driven by the assumptions on the correlation structure between α_i and x_{it}, ruled out in the RE framework and allowed for in the FE framework. In both cases, the assumption of strict exogeneity is needed for consistency, where x_{it} is assumed to be uncorrelated with e_{is} at any time $s = 1, ..., T$.[3]

Under the assumptions of orthogonality, strict exogeneity, and normally distributed errors, a maximum likelihood (ML) approach can be employed for estimation. For given σ^2_α and σ^2_e, the ML estimator for β is the same as the GLS estimator. When σ^2_α and σ^2_e have to be estimated, GLS is usually applied in two steps, while Gaussian ML could be obtained from iterating GLS to convergence.[4] Nonetheless, with normally distributed errors, ML and GLS estimators are asymptotically equivalent, and over time the latter has been preferred as computationally simpler (Balestra and Nerlove, 1966; Maddala and Mount, 1973; Cameron and Trivedi, 2005, pag.734).

If the orthogonality condition is not satisfied, the RE approach (GLS or ML) is usually claimed to be biased and inconsistent.

As a major exception, the first result of this paper shows cases where an endogenous

[1]The more general specification includes three sources of variation. Besides the individual component α_i and the idiosyncratic error term e_{it}, a time component τ_t can be considered, which varies over time and is constant across all units. As the panel analysis usually focuses on the 'heterogeneity' across individuals, in order to simplify our analysis, we will assume $\tau_t = 0$. As the time dimension is typically small, the time effects can be controlled for by including time dummies in the regression.

[2]For an alternative approach see Kmenta (1986, ch.12) who proposes a cross-sectional heteroskedastic and timewise autoregressive model, allowing both for heteroskedasticity among units and autocorrelation over time.

[3]In addition, first differences of the data can be considered, and model in (1) is estimated by ordinary least squares (OLS) regression of $y_{it} - y_{i,t-1}$ on $x_{it} - x_{i,t-1}$. Weaker assumptions are needed for the consistency of the estimator, as x_{it} is required to be uncorrelated with e_{is} with $s = t-1, t, t+1$. Still, feedback effects are ruled out.

[4]The ML and GLS estimators of the variance components differ in terms of degree of freedom adjustment (see e.g., Cameron and Trivedi, 2005, pag. 736; Hsiao, 2003, pag. 38-40).

x_{it} $(t = 1, ..., T)$, if correlated also with ε_{is} $(s \neq t)$, can lead iterative GLS to produce a consistent estimator.
This insight is particularly relevant for dynamic panel data models, where the regressors x_{it} include $y_{i,t-1}$, and $y_{i,t-1}$ includes $\varepsilon_{i,t-1}$, thus x_{it} is correlated with $\varepsilon_{i,t-1}$. Within this setting, the assumptions underlying the FE and RE frameworks for estimation are clearly violated, and an instrumental variable approach has been first proposed (Anderson, Hsiao, 1981, 1982), followed by GMM methods (Arellano, Bond, 1991; Arellano, Bover, 1995; Blundell, Bond, 1998). More recently, a transformed likelihood approach has been proposed for the estimation of linear dynamic model within a FE framework (Hsiao, Pesaran, Tahmiscioglu, 2002).
In this paper we start from a Gaussian ML approach (iterative GLS), and we take into account the probability limit of the score function. We first show why x_{it} correlated with α_i leads nevertheless to consistency, provided that x_{it} is correlated with ε_{is} $(s \neq t)$. Then we focus on the particular (but most interesting in practice) case of the dynamic model. As well known, Gaussian ML provides a consistent estimator of the autoregressive parameter in cases where the initial observation y_{i0} is unrelated with the corresponding error term. In the more interesting case where correlation is allowed for between y_{i0} and the unit heterogeneity, ML estimation is inconsistent. However the "bias" of the score can be estimated, and an "adjusted score" can be used for estimation. Furthermore, the set of moment conditions based on the "adjusted score" can be fruitfully employed within the GMM framework leading to an improved estimation efficiency. Even better, a "new single" moment condition can be obtained. Adding it to a "parsimonious" set of moment conditions (e.g. Anderson, Hsiao, 1981, 1982), it can beat the efficiency of the most adopted GMM methods, driving down the number of moment conditions, thus avoiding the well known problem in small sample caused by instrument proliferation (Roodman, 2009).
The paper proceeds as follows. The next section sets forth the basic intuition underlying our analysis in a static framework. Section 3 considers a dynamic specification and provides details on the proposed methodology. Monte Carlo experiments are considered in section 4. Section 5 concludes.

2 Static framework

We consider the estimation of the static model defined in (1).[5] The model can also be seen as a seemingly-unrelated regression model with T equations and N observations (Bhargava, Sargan, 1983). For simplicity we let the number of regressors

[5] At the onset, the presence of lagged values of y_{it} on the right hand side is ruled out. The dynamic case, i.e. the case where lagged values of the endogenous variable are included in the regression is taken into account in section 3.

k equal to 1,[6] and we define the matrix of regressors X' of size $T \times NT$:

$$X' = \left[\begin{array}{cccc|cccc|c|cccc} x_{11} & x_{21} & \dots & x_{N1} & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & x_{12} & x_{22} & \dots & x_{N2} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & x_{1T} & x_{2T} & \dots & x_{NT} \end{array}\right] \tag{2}$$

By letting $x'_t = (x_{1t}, x_{2t}, \dots, x_{Nt})$, we can write

$$X' = \left[\begin{array}{c|c|c|c} x'_1 & 0 & \dots & 0 \\ 0 & x'_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x'_T \end{array}\right] \tag{3}$$

Accordingly, we let:

$$\varepsilon' = [\varepsilon_{11}, \varepsilon_{21}, \dots, \varepsilon_{N1} | \varepsilon_{12}, \varepsilon_{22}, \dots, \varepsilon_{N2} | \dots | \varepsilon_{11}, \dots, \varepsilon_{NT}] = [\varepsilon'_1 | \varepsilon'_2 | \dots | \varepsilon'_T] \tag{4}$$

and $E(\varepsilon\varepsilon') = B \otimes I_N$, with B $(T \times T)$ positive definite.
Under the assumption of normally distributed errors as well as orthogonality and strict exogeneity, the (Gaussian) score function with respect to β can be written as:

$$\begin{aligned} \frac{1}{N}\frac{\partial \log L}{\partial \beta} &= \frac{\iota'_T X'(B^{-1} \otimes I_N)\varepsilon}{N} \\ &= \frac{\iota'_T}{N} \begin{bmatrix} x'_1 b^{11} & x'_1 b^{12} & x'_1 b^{13} & \dots & x'_1 b^{1T} \\ x'_2 b^{21} & x'_2 b^{22} & x'_2 b^{23} & \dots & x'_2 b^{2T} \\ \vdots & & & & \\ x'_T b^{T1} & x'_T b^{T2} & x'_T b^{T3} & \dots & x'_T b^{TT} \end{bmatrix} \left[\begin{array}{c} \varepsilon_1 \\ \hline \varepsilon_2 \\ \hline \vdots \\ \hline \varepsilon_T \end{array}\right] \\ &= \frac{\iota'_T}{N} \begin{bmatrix} b^{11} x'_1 \varepsilon_1 + b^{12} x'_1 \varepsilon_2 + \dots + b^{1T} x'_1 \varepsilon_T \\ b^{21} x'_2 \varepsilon_1 + b^{22} x'_2 \varepsilon_2 + \dots + b^{2T} x'_2 \varepsilon_T \\ \vdots \\ b^{T1} x'_T \varepsilon_1 + b^{T2} x'_T \varepsilon_2 + \dots + b^{TT} x'_T \varepsilon_T \end{bmatrix} \end{aligned} \tag{5}$$

which is equal to the sum of all the terms in the column vector (a task that is accomplished as we pre-multiply by ι'_T, a row vector of elements equal to 1).[7]

[6] By letting $k > 1$, complications in the formula arise. Main differences from our baseline case will be discussed when needed. Results are not affected.

[7] In case of $k > 1$ variables in X, we would need to pre-multiply by $\iota'_T \otimes I_k$.

2.1 Consistent estimation neglecting endogeneity

Contrary to conventional wisdom, we show cases where, even when regressors are "endogenous", the application of a "Gaussian maximum likelihood" that ignores endogeneity can provide a consistent estimator of β. Namely, we show that, in cases where x_{it} $(t = 1, ..., T)$ is correlated with ε_{is} $(s \neq t)$, iterating to convergence GLS can lead to a consistent estimate,[8] no matter whether x_{it} is correlated also with ε_{it} (thus endogenous) or not. Suppose that the correlation between the elements of x_2 and the corresponding elements of ε_1 is such that

$$x_2' = z_2' + \gamma_2 \varepsilon_1' \tag{6}$$

with z_2 exogenous and γ_2 a fixed constant.
The second row of the vector $X'(B^{-1} \otimes I_N)\varepsilon/N$ would be:

$$\frac{1}{N}\left[b^{21} z_2' \varepsilon_1 + \ldots + b^{2T} z_2' \varepsilon_T + b^{21} \gamma_2 \varepsilon_1' \varepsilon_1 + b^{22} \gamma_2 \varepsilon_1' \varepsilon_2 + \ldots + b^{2T} \gamma_2 \varepsilon_1' \varepsilon_T\right] \tag{7}$$

By letting $N \to \infty$ and from exogeneity of z_2, we get

$$0 + \ldots + 0 + \gamma_2 [b^{21} b_{11} + b^{22} b_{21} + \ldots + b^{2T} b_{T1}] = \gamma_2 [b^{2\bullet} b_{\bullet 1}] = 0 \tag{8}$$

as we have the product of row 2 of B^{-1} with column 1 of B.
Generally, we will get 0 if elements in x_t $(t = 1, ..., T)$ are correlated with the corresponding elements in ε_s with $s \neq t$ (analogously to equation 6), as the formula would include the product of row t of B^{-1} with column s of B. This would allow for correlation with the individual effect and idiosyncratic error term at any previous or subsequent time periods. If this pattern of correlation holds, the probability limit of the score is zero for any structure of the matrix B.[9]
This would be a case where neglected endogeneity of the regressors does not affect consistency of the ML estimator. Estimation can be performed iterating a GLS-type algorithm: iterations stop when the empirical analogue of (5) is zero.

2.2 Inconsistency: strict exogeneity without orthogonality

Consider the case where x_{it} is correlated with the individual effect α_i, but not with any e_{is} $(s = 1, ..., T)$, i.e. the standard case where the orthogonality condition is not satisfied, but we can assume strict exogeneity: $x_t' = z_t' + \gamma_t \alpha'$. Let $B =$

[8] Strictly speaking, iterative GLS is not a ML estimator in this context. A correctly specified likelihood would require the full specification of the pattern of endogeneity of x.

[9] The result holds even if matrix B departs from the standard decomposition considered in the random effect framework, i.e. $B \neq \sigma_e^2 I_T + \sigma_\alpha^2 \iota_T \iota_T'$.

$\sigma_e^2 I_T + \sigma_\alpha^2 \iota_T \iota_T'$ and $\mu = \sigma_e^2/\sigma_\alpha^2$, then the probability limit of the t-th element of the vector on the r.h.s. of equation (5) is:

$$\gamma_t[b^{t1}\sigma_\alpha^2 + b^{t2}\sigma_\alpha^2 + \ldots + b^{tT}\sigma_\alpha^2] = \frac{\gamma_t}{\mu}\left[1 - \frac{T}{T+\mu}\right] = \frac{\gamma_t}{T+\mu} \neq 0 \tag{9}$$

unless $\gamma_t = 0$, i.e. the orthogonality condition holds.

2.3 Consistency in a special instance of simultaneity

In this case we take into account a very special case (presumably not of interest in practice; but it might happen!). Iterating to convergence GLS can also provide consistent estimates in the case when x_t is correlated not only with α but also with e_t, provided it has a particular negative correlation with e_t; that is $x_t' = z_t' + \gamma_t \alpha' - \gamma_t \frac{1}{\sigma_e^2(T+\mu)} e_t'$, where $\mu = \sigma_e^2/\sigma_\alpha^2$. If the matrix B has the "classical" equicorrelated structure, i.e. $B = \sigma_e^2 I_T + \sigma_\alpha^2 \iota_T \iota_T'$, then row t of the right hand side of equation (5) has the following probability limit ($N \to \infty$):

$$\gamma_t\left[b^{t1}\sigma_\alpha^2 + b^{t2}\sigma_\alpha^2 + \ldots + b^{tT}\sigma_\alpha^2 - \frac{\sigma_e^2}{\sigma_e^2(T+\mu)}\right] = \gamma_t\left[\frac{1}{T+\mu}\right] - \gamma_t\left[\frac{1}{T+\mu}\right] = 0 \tag{10}$$

3 Dynamic model

As a particular case, let us consider $x_t = y_{t-1}$:[10]

$$y_{it} = \beta y_{it-1} + \varepsilon_{it} = \beta y_{it-1} + \alpha_i + e_{it} \tag{11}$$

where we assume that y_{i0} is observed ($T+1$ observations are available for estimation). The observation $x_t = y_{t-1}$ "contains" ε_{it-1}, and thus α_i; it is therefore an endogenous regressor, at the same time correlated with an ε_{is} ($s \neq t$).
The score function in (5) (Gaussian score ignoring endogeneity) becomes:

$$\frac{1}{N}\frac{\partial \log L}{\partial \beta} = \frac{\iota_T'}{N}\begin{bmatrix} b^{11}y_0'\varepsilon_1 + b^{12}y_0'\varepsilon_2 + \ldots + b^{1T}y_0'\varepsilon_T \\ b^{21}y_1'\varepsilon_1 + b^{22}y_1'\varepsilon_2 + \ldots + b^{2T}y_1'\varepsilon_T \\ \vdots \\ b^{T1}y_{T-1}'\varepsilon_1 + b^{T2}y_{T-1}'\varepsilon_2 + \ldots + b^{TT}y_{T-1}'\varepsilon_T \end{bmatrix} \tag{12}$$

If y_{i0} is exogenous both with respect to α_i and e_{i0}, then $\frac{1}{N}[b^{11}y_0'\varepsilon_1 + b^{12}y_0'\varepsilon_2 + \ldots + b^{1T}y_0'\varepsilon_T] \xrightarrow{p} 0$. Also the other terms of the score (i.e. the rows on the right hand

[10] In the case of k regressors, y_{t-1} is one of them and the other $k-1$ variables are exogenous.

side of equation 12) have a zero probability limit (as it was shown in section 2.1). This is well known in the literature: Gaussian maximum likelihood estimation does provide a consistent estimator of β with exogenous initial conditions (Anderson, Hsiao, 1982; Bhargava, Sargan, 1983). Furthermore, by writing the dynamic model as a system of T equations over N individuals we get a triangular structure; therefore by iterating to convergence a GLS-type estimator we get a consistent and efficient estimator even if the regressors y_{i1},...,$y_{i,T-1}$ are endogenous (see Lahiri and Schmidt, 1978, extended to the case of simultaneous equations by Calzolari and Sampoli, 1993).
More interestingly, if y_0 is correlated with α, the rough (Gaussian) maximum likelihood does not provide a consistent estimator.
Estimation of this model largely relies on GMM. Two main frameworks are employed in the empirical analysis: GMM-difference estimator, proposed by Arellano, Bond (1991), and the GMM-system estimator (Arellano, Bover, 1995; Blundell, Bond, 1998).[11]
The GMM-difference estimator first transforms the model using first differences to remove the individual effect α_i, and then, under the assumption of uncorrelated e_{it}, uses lag 2 and older of y_{it} as instruments for Δy_{it-1}. The set of moment conditions used for GMM estimation can be written as $(j, t \geq 2)$:

$$E[y_{it-j}\Delta\varepsilon_{it}] = 0 \tag{13}$$

As y_{i0} is observed, we get $T(T-1)/2$ moment conditions. Note that under the assumption that e_{it} is uncorrelated over time, the covariance between the initial condition y_{i0} and the composite error term is only driven by the presence of the individual effect α_i and is therefore constant over time, i.e. we can write:

$$E[y_{i0}\varepsilon_{it}] = E[y_{i0}\alpha_i] = \sigma_{0\varepsilon} \text{ for each } t \geq 1 \tag{14}$$

The GMM-system estimator introduces an additional hypothesis by adding the following $T-1$ moment conditions $(t \geq 2)$ to the moment conditions in (13):

$$E[\Delta y_{it-1}\varepsilon_{it}] = 0 \tag{15}$$

Despite this additional restrictive assumption, the GMM-system method is widely used due to its superior performance with respect to the GMM-difference estimator, especially as the value of β approaches 1 (Blundell, Bond, 1998).
We will now show the relationship between conditions (15) and (14), and explain why (15) is more restrictive. We shall then show how the less restrictive condition (14) is exploited in the set up we propose.

[11] Please refer to the original papers for a detailed description. In the following we will only focus on the characteristics of the two methods that are relevant for comparison with our estimation strategy.

By letting $t = 2$, equation (15) becomes:

$$\begin{aligned} 0 &= E[\Delta y_{i1}\varepsilon_{i2}] = E[(y_{i1} - y_{i0})\varepsilon_{i2}] \\ &= E[(\beta y_{i0} + \varepsilon_{i1} - y_{i0})\varepsilon_{i2}] \\ &= -(1-\beta)E[y_{i0}\varepsilon_{i2}] + E[\varepsilon_{i1}\varepsilon_{i2}] \end{aligned}$$

As (for the lack of serial correlation in e_{it}) $E[\varepsilon_{i1}\varepsilon_{i2}] = \sigma^2_\alpha$, this is equivalent to:

$$E[y_{i0}\varepsilon_{i2}] = \frac{\sigma^2_\alpha}{1-\beta}$$

In general terms, by recurrent substitution, the moment conditions in (15) can be written as ($t \geq 3$):

$$\begin{aligned} 0 &= E[\Delta y_{it-1}\varepsilon_{it}] = E[(y_{it-1} - y_{it-2})\varepsilon_{it}] \\ &= E\left[\left((\beta-1)\beta^{t-2}y_{i0} + (\beta-1)\sum_{j=0}^{t-3}\beta^j\varepsilon_{it-j-2} + \varepsilon_{it-1}\right)\varepsilon_{it}\right] \\ &= (\beta-1)\beta^{t-2}E[y_{i0}\varepsilon_{it}] + \sum_{j=0}^{t-3}(\beta^{j+1} - \beta^j)E[\varepsilon_{it-j-2}\varepsilon_{it}] + E[\varepsilon_{it-1}\varepsilon_{it}] \\ &= (\beta-1)\beta^{t-2}E[y_{i0}\varepsilon_{it}] + \sum_{j=0}^{t-3}(\beta^{j+1} - \beta^j)\sigma^2_\alpha + \sigma^2_\alpha \\ &= (\beta-1)\beta^{t-2}E[y_{i0}\varepsilon_{it}] + (\beta^{t-3+1} - \beta^0)\sigma^2_\alpha + \sigma^2_\alpha \\ &= (\beta-1)\beta^{t-2}E[y_{i0}\varepsilon_{it}] + \beta^{t-2}\sigma^2_\alpha \end{aligned}$$

that is:

$$0 = (\beta-1)E[y_{i0}\varepsilon_{it}] + \sigma^2_\alpha$$

Summing up, condition (15) can also be written as

$$E[y_{i0}\varepsilon_{it}] = \frac{\sigma^2_\alpha}{1-\beta} \quad \text{for each } t \geq 2 \tag{16}$$

a condition that is more stringent than (14). In fact, while equation (14) simply assumes that covariance between y_{i0} and the corresponding α_i is constant for any i, equation (16) also implies that y_{i0} have been produced by the same "stationary in α" process that we assume for $t \geq 1$ (equation 11).
Recent literature has stressed the fact that the number of instruments to be used with GMM-difference and GMM-system estimators fast increases with T and has

highlighted their poor performance when instruments are "too many" (Roodman, 2009; Ziliak, 1997). By building on the insights of section 2.1 and exploiting the information in (14), in this paper we propose an estimation strategy that allows to avoid the problem of instrument proliferation with gains in efficiency. The covariance $\sigma_{0\varepsilon}$ is estimated and treated as a nuisance parameter.
Consider the first part of the score ($t = 1$, corresponding to the first row on the r.h.s. 12). This can be written as:

$$\begin{aligned}\frac{1}{N}[b^{11}y_0'\varepsilon_1 + b^{12}y_0'\varepsilon_2 + \ldots + b^{1T}y_0'\varepsilon_T] &\xrightarrow{p} b^{11}E[y_0'\varepsilon_1] + b^{12}E[y_0'\varepsilon_2] + \ldots + b^{1T}E[y_0'\varepsilon_T] \\ &= \sigma_{0\varepsilon}\left[b^{11} + b^{12} + ... + b^{1T}\right] \qquad (17)\end{aligned}$$

where $E[y_{i0}\varepsilon_{it}] = \sigma_{0\varepsilon}$ as in equation (14). The probability limit is different from zero, as the values in B take into account times from 1 to T, whereas correlation is considered with error terms that do not enter into the construction of B (i.e. $t = 0$).
For $t = 2$, we have:

$$\begin{aligned}\frac{1}{N}[b^{21}y_1'\varepsilon_1 + b^{22}y_1'\varepsilon_2 + \ldots + b^{2T}y_1'\varepsilon_T] &= \frac{1}{N}[b^{21}(\beta y_0 + \varepsilon_1)'\varepsilon_1 + b^{22}(\beta y_0 + \varepsilon_1)'\varepsilon_2 \\ &\quad + \ldots + b^{2T}(\beta y_0 + \varepsilon_1)'\varepsilon_T] \\ &= \frac{1}{N}\beta[b^{21}y_0'\varepsilon_1 + b^{22}y_0'\varepsilon_2 + \ldots + b^{2T}y_0'\varepsilon_T] \\ &\quad + \frac{1}{N}[b^{21}\varepsilon_1'\varepsilon_1 + b^{22}\varepsilon_1'\varepsilon_2 + \ldots + b^{2T}\varepsilon_1'\varepsilon_T] \\ &\xrightarrow{p} \beta\sigma_{0\varepsilon}\left[b^{11} + b^{12} + ... + b^{1T}\right] + b^{2\bullet}b_{\bullet 1} \\ &= \beta\sigma_{0\varepsilon}\left[b^{11} + b^{12} + ... + b^{1T}\right] \qquad (18)\end{aligned}$$

as $b^{2\bullet}b_{\bullet 1} = 0$.
In general terms we have ($t = 1, ..., T$):

$$\frac{1}{N}[b^{t1}y_{t-1}'\varepsilon_1 + b^{t2}y_{t-1}'\varepsilon_2 + \ldots + b^{tT}y_{t-1}'\varepsilon_T] \xrightarrow{p} \beta^{t-1}\sigma_{0\varepsilon}\left[b^{t1} + b^{t2} + ... + b^{tT}\right] \qquad (19)$$

Put it differently:

$$\frac{1}{N}[b^{t1}y_{t-1}'\varepsilon_1 + b^{t2}y_{t-1}'\varepsilon_2 + \ldots + b^{tT}y_{t-1}'\varepsilon_T] - \beta^{t-1}\sigma_{0\varepsilon}\left[b^{t1} + b^{t2} + ... + b^{tT}\right] \xrightarrow{p} 0 \qquad (20)$$

These equations might suggest a set of T orthogonality conditions to be exploited in a GMM framework. However, in order to reduce the number of moment conditions, we propose to sum up the (empirical analogues of the) T moment equations in (20) over $t = 1, ..., T$, obtaining a single "new" moment condition (to be set equal to zero):

$$\frac{1}{N}\begin{bmatrix} b^{11}y_0'\varepsilon_1 + ... + b^{1T}y_0'\varepsilon_T \\ +b^{21}y_1'\varepsilon_1 + ... + b^{2T}y_1'\varepsilon_T \\ \vdots \\ +b^{T1}y_{T-1}'\varepsilon_1 + ... + b^{TT}y_{T-1}'\varepsilon_T \end{bmatrix} - \begin{bmatrix} \sigma_{0\varepsilon}[b^{11} + ... + b^{1T}] \\ +\beta\sigma_{0\varepsilon}[b^{21} + ... + b^{2T}] \\ \vdots \\ +\beta^{T-1}\sigma_{0\varepsilon}[b^{T1} + ... + b^{TT}] \end{bmatrix} \tag{21}$$

In this paper we propose to add this single moment condition to existing GMM estimators that do not use "too many" instruments. As it will be shown in the next section, we may obtain, in cases of practical interest, a triple benefit:
(1) improved efficiency (verified by Monte Carlo experiments);
(2) reduction of the number of moment conditions over the most adopted GMM methods;
(3) wider applicability, as y_{i0} do not need to be "stationary endogenous".
During GMM iterative procedure, the last iteration residuals are used to estimate the covariance matrix B; together with the observed y_{i0} they also produce the estimate of $\sigma_{0\varepsilon}$ to be used in the next iteration.

4 Monte Carlo experiment

In this section we explore the performance of the proposed approach.[12] We generated a dynamic panel data model as:

$$y_{it} = \beta y_{i,t-1} + \varepsilon_{it} = \beta y_{i,t-1} + \alpha_i + e_{it} \tag{22}$$

where $|\beta| < 1$, $\alpha_i \sim N(0, 1)$, $e_{it} = \delta_i\tau_t w_{it}$ with $\delta_i \sim U(0.5, 1.5)$, $\tau_t = 0.5 + 0.1(t-1)$ (for $t > 0$), and $w_{it} \sim \chi^2(1) - 1$. As a result, the idiosyncratic component e_{it} exhibits heteroschedasticity both over time and across units, as well as asymmetry. A sort of "stationarity for α" is produced generating a fifty-time-period pre-sample, with $\tau_t = 0.5$ for $t = -50, ..., 0$ before the estimation sample is drawn up to $t = T$.[13] We assume that y_{i0} is observed, therefore $T + 1$ time periods are included in the data. We also consider a "non-stationary case", where the pre-sample draws are still considered, but y_{i0} is generated as exogenous with respect to α_i (we keep the variance of y_{i0} unchanged with respect to the previous case).
Table 1 reports the results of the Monte Carlo experiments (10,000 replications) with $\beta = 0.1, 0.5, 0.9$ and $T = 5$ (with y_{i0} observed). We start from the simple IV

[12]Simulation experiments are run using Fortran 77.
[13]The generation of e_{it} closely resembles the generating process employed by Windmeijer (2000, p. 34).

estimator proposed by Anderson, Hsiao (1981, 1982), IV-AH, where first differences of the original model are considered for estimation (in order to remove the individual heterogeneity α_i), and $y_{i,t-2}$ is taken as an instrument for $\Delta y_{i,t-1}$; 4 time periods are thus available for estimation and 4 instruments are used. The same set of instruments is then used within a GMM framework, and will be labeled GMM-AH. This is one of the most "parsimonious" among the methods proposed in the literature in terms of the number of employed moment conditions ("only" $T-1$). The proposed moment condition (21) is then added to this set of instruments and the method, labeled GMM-AH-CM (AH and Calzolari-Magazzini), employes T moment conditions.

More commonly used by practitioners, the GMM-difference estimator proposed by Arellano, Bond (1991) exploits not only lag 2 but also all previous lags of y_{it} as instruments for the equation in first differences (GMM-AB). The moment condition (21) is also added here: GMM-AB-CM. As widely acknowledged by the literature, the number of moment conditions employed for estimation by GMM-AB increases quadratically with T, posing concerns about the performance of GMM estimation when instruments are "too many" (see Roodman, 2009 for a review of available evidence).

The Monte Carlo mean of the estimated coefficient is "close to the true value", with the exception of the AH (GMM and IV) estimator when $\beta = 0.9$ with endogenous y_0.[14] As also highlighted by Arellano, Bond (1991), lack of identification seems to arise when using AH. Note that in this case ($\beta = 0.9$, y_0 endogenous) when the proposed moment condition is added to the AH set (GMM-AH-CM), the performance of the estimator increases and the average of Monte Carlo coefficients gets much closer to the true value.

This is also the case with the GMM-AB estimator where the additional moment condition (GMM-AB-CM) leads to a slight decrease in bias and to a substantial reduction in the variance of the estimator (from .86E-1 to .36E-1). This is not the case for smaller values of β, where the additional moment condition (21) does not bring significant gains to the performance of both the (GMM) AH and AB estimators.

In the cases where y_0 is exogenous, adding the condition (21) improves the per-

[14] The GMM-system estimator (Blundell, Bond, 1998) is not considered here as it would not be consistent in the case where y_0 is exogenous or endogenous but not mean stationary. Unreported Monte Carlo analysis performed using the STATA command `xtdpdsys` show that the GMM-system estimator outperforms the proposed approach (GMM-AH-CM) only with endogenous y_0 and no constant term in the model. When an intercept is added to the data generating process (see Table 2) the performance of the GMM-AH-CM estimator is comparable to the GMM-system approach with the advantage of a reduced set of moment conditions (5 versus 14 in this setting) and robustness to the lack of mean stationarity hypothesis required for consistency of the GMM-system. The full set of results is available from the authors upon request.

	IV AH	GMM AH	GMM AH-CM	GMM AB	GMM AB-CM
			y_0 endogenous		
$\beta = 0.1$					
Mean	.1000	.1007	.1003	.1019	.1002
Variance	.17E-2	.15E-2	.15E-2	.14E-2	.14E-2
$\beta = 0.5$					
Mean	.4998	.4986	.4991	.5047	.5069
Variance	.54E-2	.45E-02	.46E-2	.37E-2	.37E-2
$\beta = 0.9$					
Mean	.6433	42.85	.8966	.8862	.9115
Variance	476.0	.23E+7	.1632	.86E-1	.36E-1
			y_0 exogenous		
$\beta = 0.1$					
Mean	.1000	.1004	.1004	.1005	.0996
Variance	.98E-3	.69E-3	.64E-3	.44E-3	.45E-3
$\beta = 0.5$					
Mean	.4999	.4997	.5007	.4996	.4997
Variance	.16E-2	.82E-3	.50E-3	.29E-3	.29E-3
$\beta = 0.9$					
Mean	.8998	.8997	.9000	.8996	.8996
Variance	.26E-3	.18E-3	.18E-3	.11E-3	.11E-3
N. instruments	4	4	5	10	11

Table 1: Monte Carlo results (10,000 replications), $N = 1000$, $T = 5$ (y_0 observed)

formance of AH except when $\beta = 0.9$, whereas the performance of AB is basically unchanged.
The results of further experiments with different instances of "non-stationarity" are reported in Table 2. Starting from our baseline data generating process (22),

(i) we added a constant term in the generation of y_{i0} only, that is y_{i0} is generated as $y_{i0} = \mu + \beta y_{i,-1} + \varepsilon_{i0}$ and equation (22) is considered for $t \neq 0$ (pre-sample draws as described before are still considered);

(ii) we added a constant term to the pre-sample draws but not to the estimation data ($t = 1, ..., T$), that is $y_{it} = \mu + \beta y_{i,t-1} + \varepsilon_{it}$ for $t \leq 0$, while equation (22) is considered for $t > 0$.

We set $\mu = 10$ in all the experiments.[15]
The identification problems with AH in Table 1 now disappears. All estimators are consistent with significant gains spanning from the application of the proposed methodology. Adding the one moment equation defined in (21) leads to significant gains for the GMM-AH estimator and makes the additional moment conditions employed by AB (i.e. lags 3 and beyond of y as instruments for the differenced equation) redundant. The use of the additional moment condition (21) significantly improves on the performance of the AH estimator leading to a decrease in the variance of about 40% with $\beta = 0.1$ and $\beta = 0.5$, up to 90% with $\beta = 0.9$ in the endogenous cases (79% when y_0 is exogenous). Even though GMM-AB is more efficient than GMM-AH, the method is less efficient than the GMM-AH-CM estimator (that also matches the variance of the GMM-AB-CM estimator). Furthermore, GMM-AH-CM allows to keep the number of moment conditions remarkably small.

5 Summary

Starting from a (Gaussian) maximum likelihood approach for the estimation of linear panel data models, this paper takes into account the convergence conditions of iterative GLS. We have considered models where neither the orthogonality condition nor the strict exogeneity requirement are satisfied, and we show interesting cases where, iterating to convergence a GLS estimator neglecting the endogeneity

[15] In case (ii) the long run average (the value met in y_{i0}) is $\mu/(1-\beta)$, i.e. it is increasing with the value of β. Set up (i) and (ii) produce similar results when the same value is met in y_{i0}. As an example, if we let $\mu = 5$ when $\beta = 0.5$ in (ii) with $\mu/(1-\beta) = 10$, the performance of the proposed estimator is identical to the case where $\mu = 10$ (and $\beta = 0.5$) in (i). The same result emerges if we set $\mu = 1$ when $\beta = 0.9$ in (ii) (where $1/(1-0.9) = 10$) as compared to setting $\mu = 10$ in (i).

	IV AH	GMM AH	GMM AH-CM	GMM AB	GMM AB-CM
	intercept added in the pre-sample, y_0 endogenous				
$\beta = 0.1$					
Mean	.1000	.1003	.1004	.1005	.1004
Variance	.12E-4	.11E-4	.68E-5	.75E-5	.69E-5
$\beta = 0.5$					
Mean	.5000	.5002	.5001	.5003	.5002
Variance	.52E-5	.40E-5	.26E-5	.35E-5	.27E-5
$\beta = 0.9$					
Mean	.9000	.9000	.9000	.9002	.8999
Variance	.17E-5	.14E-5	.14E-6	.14E-5	.14E-6
	intercept added in $t = 0$, y_0 endogenous				
$\beta = 0.1$					
Mean	.1000	.1003	.1004	.1006	.1005
Variance	.15E-4	.14E-4	.84E-5	.92E-5	.85E-5
$\beta = 0.5$					
Mean	.5000	.5003	.5002	.5007	.5004
Variance	.23E-4	.19E-4	.11E-4	.15E-4	.11E-4
$\beta = 0.9$					
Mean	.8997	.9008	.8988	.9010	.8990
Variance	.42E-3	.29E-3	.27E-4	.23E-3	.26E-4
	intercept added in $t = 0$, y_0 exogenous				
$\beta = 0.1$					
Mean	.1000	.1003	.1004	.1006	.1006
Variance	.15E-4	.14E-4	.82E-5	.90E-5	.83E-5
$\beta = 0.5$					
Mean	.5000	.5003	.5003	.5007	.5005
Variance	.22E-4	.18E-4	.10E-4	.13E-4	.10E-4
$\beta = 0.9$					
Mean	.8998	.9004	.8998	.9008	.9000
Variance	.10E-3	.78E-4	.16E-4	.59E-4	.16E-4
N. instruments	4	4	5	10	11

Table 2: Monte Carlo results (10,000 replications), $N = 1000$, $T = 5$ (y_0 observed), experiments with intercept

of the explanatory variables, can lead to a consistent estimator. Namely, if x_t is related to ε_s $(t \neq s)$, we get a consistent estimator of the parameter of interest (β).
Since this is the typical condition when $x_t = y_{t-1}$ (correlated with ε_{t-1}), we have focused on the dynamic case. By developing the "score function" that neglects endogeneity of y_{t-1}, we propose a new moment condition that can be employed for GMM estimation. Monte Carlo simulations show that the proposed approach has three advantages: (i) it avoids the problem of instrument proliferation by adding a single moment condition to the "parsimonious" set of instruments proposed by Anderson, Hsiao (1981, 1982), leading to "only T" moment conditions for estimation; (ii) improvements in performance with respect to standard GMM methods can be substantial; (iii) it does not require additional assumptions with respect to Anderson, Hsiao (1981, 1982), and Arellano, Bond (1991).

References

[1] Anderson, T. and Hsiao, C.: 1981, "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association* **76**, 598–606.

[2] Anderson, T. and Hsiao, C.: 1982, "Formulation and Estimation of Dynamic Models Using Panel Data", *Journal of Econometrics* **18**, 47–82.

[3] Arellano, M. and Bond, S.: 1991, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *The Review of Economic Studies* **58**(2), 277–297.

[4] Arellano, M. and Bover, O.: 1995, "Another Look at the Instrumental Variables Estimation of Error Components Models', *Journal of Econometrics* **68**, 29-51.

[5] Balestra, P. and Nerlove, M.: 1966, "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas", *Econometrica* 34(3), 585–612.

[6] Bhargava, A. and Sargan, J.D.: 1983, "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods", *Econometrica* **51**(6), 1635–1659.

[7] Blundell, S. and Bond, S.: "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics* **87**(1), 115–143.

[8] Calzolari, G. and Sampoli, L.: 1993, "A Curious Result on Exact FIML and Instrumental Variables", *Econometric Theory* **9**, 296–309.

[9] Cameron, A.C. and Trivedi, P.K.: 2005, *Microeconometrics*, Cambridge University Press.

[10] Hansen, L.: 1982, "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica* **50**, 1029-1054.

[11] Hsiao, C.: 2003, *Analysis of Panel Data*, Cambridge University Press.

[12] Hsiao, C., Pesaran, M.H. and Tahmiscioglu A.K.: 2002, "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods", *Journal of Econometrics* **109**(1), 107–150.

[13] Kmenta, J.: 1986, *Elements of Econometrics* (2nd ed.), New York: Macmillan Publishing Company.

[14] Lahiri, K. and Schmidt, P.: 1978, "On the Estimation of Triangular Structural Systems", *Econometrica* **46**(5), 1217–1221.

[15] Maddala, G. and Mount, T.: 1973, "A Comparative Study of Alternative Estimators for Variance Components Models Used in Econometric Applications", *Journal of the American Statistical Association* **68**(342), 324–328.

[16] Mundlak, Y.: 1978, "On the Pooling of Time Series and Cross Section Data", *Econometrica* **46**(1), 69–85.

[17] Roodman, D.: 2009, "A Note on the Theme of Too Many Instruments", *Oxford Bulletin of Economics and Statistics* **71**(1), 135–158.

[18] Windmeijer, F.: 2000, "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators", *Journal of Econometrics* **126**(1), 25–51.

[19] Ziliak, J.P.: 1997, "Efficient Estimation with Panel Data When Instruments are Predetermined: An Empirical Comparison of Moment-Condition Estimators", *Journal of Business and Economic Statistics* **15**(4), 419–431.