Working Paper Series
Department of Economics
University of Verona

Identification of linear panel data models when instruments are not available

Laura Magazzini, Giorgio Calzolari

WP Number: 6 February 2012

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

Identification of linear panel data models when instruments are not available

Laura Magazzini* Giorgio Calzolari†

February 13, 2012

Abstract

One of the major virtues of panel data models is the possibility to control for unobserved and unobservable heterogeneity at the unit (individual, firm, sector...) level, even when this is correlated with the variables included on the right hand side of the equation. By assuming an additive error structure, identification of the model parameters spans from transformations of the data that wipe out the individual component. We propose an alternative identification strategy, where the equation of interest is embedded in a structural system that properly accounts for the endogeneity of the variables on the right hand side (without distinguishing correlation with the individual component or the idiosyncratic term). We show that, under certain conditions, the system is identified even in the case where no exogenous variable is available, due to the presence of cross-equation restrictions. Estimation of the model parameters can rely on an iterated Zellner-type estimator, with remarkable performance gains over traditional GMM approaches.
Keywords: panel data, identification, cross-equation restrictions.
JEL Codes: C23, C33.

1 Introduction

In this paper we propose a new approach for the identification of linear panel data models of N units (individuals, firms, countries, regions, etc.) observed over $T \geq 2$ time periods:[1]

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

The error term ε_{it} is usually decomposed into two independent sources: $\alpha_i + e_{it}$, in which α_i captures unit 'unobserved heterogeneity', constant over time and different across units, and e_{it} is an idiosyncratic component varying over time and across units. We focus on cases in which the exogeneity of x_{it} is ruled out.[2] One of the major virtues of panel data is that they allow the identification of β even in cases when the regressors x_{it} are correlated with the error term. Arellano and Honoré (2001) discusses the identification strategies of linear panel data

*Department of Economics, Università di Verona, <laura.magazzini@univr.it>
†Department of Statistics, Università di Firenze, <calzolar@ds.unifi.it>

[1] For simplicity a balanced panel dataset is considered. No differences in results appear with unbalanced panel with no attrition, where a different number of observation, T_i, is available (at random) for each unit. Our focus is on the case where N is large and T is small.

[2] Put it differently, we set aside the early distinction between the fixed effect and random effect approaches (Mundlak, 1978), and focus the discussion on the identification of β in cases in which x_{it} is correlated with the error term (α_i or e_{it} – or its past values – or both).

models under various assumptions on the structure of correlation. In the simple case in which x_{it} is correlated with α_i, identification is achieved imposing a "strict exogeneity" assumption, i.e. $E^*(e_{it}|x_{i1}, x_{i2}, ...x_{iT}) = 0$ (with E^* denoting a linear projection), and considering a transformation of the data that removes the α_i from the equation (either within group or first difference transformation). The strict exogeneity assumption also assures identification of partial adjustment models when $T \geq 3$ (Arellano and Honoré, 2001). If x_{it} is treated as a predetermined variable, identification spans from restrictions on the correlation structure of e_{it}, namely, lack of autocorrelation is imposed on e_{it} (Arellano and Honoré, 2001; Anderson and Hsiao, 1981; Anderson and Hsiao, 1982; Arellano and Bond, 1991; Ahn and Schmidt, 1995 – see also Arellano and Bover, 1995; Blundell and Bond, 1998).
In this paper we propose an alternative strategy that treats the linear panel data model in (1) as a system of T equations over N observations (Bhargava and Sargan, 1983). In Section 2 we show that, under certain conditions, the system is identified even in the case when no exogenous variable is available (besides the constant term). Identification is achieved trough cross-equation restrictions, spanning from the assumption of time-invariant coefficients in the equation of interest (Wegge, 1965; Kelly, 1975). In Section 3 a Monte Carlo experiment is undertaken to show that, when the detected identification conditions are met, a general pattern of endogeneity is allowed for between the regressors included in the equation and the error term (both the idiosyncratic component α_i, and the individual component e_{it}), and estimation can rely on iterated Zellner-type estimation. Even though the discussion in the paper focuses on the static framework, the identification conditions are easily generalized to dynamic models.

2 Identification without exogenous variables

Let us consider the static linear panel data model of equation (1): $y_{it} = x'_{it}\beta + \varepsilon_{it}$, with $i = 1, ..., N$ and $t = 1, ..., T$.[3] This can be written as a system of T equations and N observations, in which coefficients are assumed to be constant over time and across units (Bhargava and Sargan, 1983).

$$\begin{cases} y_{i1} = x'_{i1}\beta + \varepsilon_{i1} \\ y_{i2} = x'_{i2}\beta + \varepsilon_{i2} \\ \cdots \\ y_{iT} = x'_{iT}\beta + \varepsilon_{iT} \end{cases} \tag{2}$$

If x is exogenous, i.e. there is no *contemporaneous* correlation between x_{it} and ε_{it}, the vector of coefficient β is identified, and estimation can be easily performed using standard methodology.
More interestingly, consider the case when the right hand side variable is endogenous, i.e. x_{it} and ε_{it} are allowed to be correlated.[4] The system in (2) is no longer identified, unless a set of identifying exogenous variables exists. In case those variables exist,[5] we can write the full structural system of equations as a system of "simultaneous" equations, simultaneity meaning "same individual", rather than "same time". In order to simplify the notation consider the case when x_{it} is a scalar and we allow the model to contain an intercept.[6] We change standard

[3]Even though the discussion focuses on the static framework, the proposed strategy is easily generalized to dynamic models, i.e. models in which y_{it-1} is included in x_{it}.
[4]Let the reader think of ε_{it} as $\varepsilon_{it} = \alpha_i + e_{it}$, as it is customarily assumed in panel data applications. Note that we allow x_{it} to be correlated both with α_i and e_{is} for some $s = 1, ..., T$. Furthermore, (even if this is never discussed in panel data applications) we can relax the assumption that the decomposition of ε_{it} into α_i and e_{it} is unique, not imposing lack of correlation between the two error terms (Berzeg, 1979).
[5]Unlikely in practice, but the discussion is still useful for our proposal.
[6]The discussion does not change if a set of endogenous variables is considered. Main differences from our baseline case will be discussed when needed.

notation in the panel literature in order to resemble the notation that is used in simultaneous equation systems, i.e. all coefficients are denoted as β_c with c a progressive number ($c = 1, ..., C$ with C the total number of coefficients in the system), and all error terms are denoted as u_{iE} where E is the number of the corresponding equation (as we have $2T$ endogenous variables, $E = 1, ..., 2T$; $i = 1, ..., N$, number of units in the panel dataset):

$$\begin{cases} x_{i1} &= \beta_1 + z_{i1}\beta_2 + u_{i1} \\ x_{i2} &= \beta_3 + z_{i2}\beta_4 + u_{i2} \\ \vdots & \\ x_{iT} &= \beta_{2T-1} + z_{iT}\beta_{2T} + u_{iT} \\ y_{i1} &= \beta_{2T+1} + x_{i1}\beta_{2T+2} + u_{i,T+1} \\ \vdots & \\ y_{iT} &= \beta_{2T+2T-1} + x_{iT}\beta_{2T+2T} + u_{i,2T} \end{cases} \tag{3}$$

We assume that z_{it} is exogenous,[7] therefore $E[z'_{it}u_{it}] = 0$ for $t = 1, ..., T$; whereas any pattern of correlation is allowed between the error terms, as well as between x_{it} and the error terms.
As the coefficient of the relationship between y_{it} and x_{it} is constant over time, the following restriction holds: $\beta_{(2T+2\tau)} = \beta$ for each $\tau = 1, 2, ..., T$. If also the intercept in the regression of y_{it} on x_{it} is not allowed to vary over time, we can add the following restriction: $\beta_{(2T+2\tau-1)} = \beta_0$ for each $\tau = 1, 2, ..., T$.
The interest is in the conditions that allow us to identify the relationship between y_i and x_i, i.e. the parameter β, when z is not observed. This is likely to be the case, because the availability of external instruments is severely limited in empirical applications. We show that, if further (reasonable in some cases) restrictions are imposed on the system parameters, identification of β does not rely on availability of the external instruments z. β can be identified even if z is not available.
Let us start with the case $T = 2$.[8] The *full* system can be written as:

$$\begin{cases} x_{i1} &= \beta_1 + z_{i1}\beta_2 + u_{i1} \\ x_{i2} &= \beta_3 + z_{i2}\beta_4 + u_{i2} \\ y_{i1} &= \beta_5 + x_{i1}\beta_6 + u_{i3} \\ y_{i2} &= \beta_7 + x_{i2}\beta_8 + u_{i4} \end{cases} \tag{4}$$

In matrix notation we have:

$$BY_i + CZ_i = U_i \tag{5}$$

with $Y_i = (x_{i1}, x_{i2}, y_{i1}, y_{i2})'$, $Z_i = (1, z_{i1}, z_{i2})'$, $U_i = (u_{i1}, u_{i2}, u_{i3}, u_{i4})'$,

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\beta_6 & 0 & 1 & 0 \\ 0 & -\beta_8 & 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} -\beta_1 & -\beta_2 & 0 \\ -\beta_3 & 0 & -\beta_4 \\ -\beta_5 & 0 & 0 \\ -\beta_7 & 0 & 0 \end{pmatrix} \tag{6}$$

with $\beta_6 = \beta_8 = \beta$. If time effects are ruled out, we can also write $\beta_5 = \beta_7 = \beta_0$.
We ask under which conditions we can identify β when z is not available, i.e. under which conditions we can recover the value of β from the reduced form coefficients in the case $Z_i = 1$, i.e. when observations about z_{i1} and z_{i2} are not available (and therefore (β_1, β_2) and (β_3, β_4) are not separately identifiable).

[7]Without loss of generality assume that z_i have mean zero.
[8]General conditions will be set forth in the Section 2.1.

Put it differently, we ask under which conditions it is possible to identify the coefficients β_6 and β_8 (where $\beta_6 = \beta_8 = \beta$) in the following system:

$$\begin{cases} x_{i1} &= \beta_1 + \tilde{u}_{i1} \\ x_{i2} &= \beta_3 + \tilde{u}_{i2} \\ y_{i1} &= \beta_5 + x_{i1}\beta_6 + u_{i3} \\ y_{i2} &= \beta_7 + x_{i2}\beta_8 + u_{i4} \end{cases} \tag{7}$$

with $\tilde{u}_{i1} = z_{i1}\beta_2 + u_{i1}$ and $\tilde{u}_{i2} = z_{i2}\beta_4 + u_{i2}$. Without further restrictions the system (7), is not identified. However, as cross-equations constraints are available, these could aid identification (Wegge, 1965; Kelly, 1975).

In matrix form, the system (7) can be written as $BY_i + C_{\cdot 1}Z_i = U_i$, with $C_{\cdot 1}$ denoting the first column of C and $Z_i = 1$ (all the other matrices and vectors as previously defined). The reduced form parameters can be obtained as:

$$\Pi_{\cdot 1} = B^{-1}C_{\cdot 1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \beta_6 & 0 & 1 & 0 \\ 0 & \beta_8 & 0 & 1 \end{pmatrix} \begin{pmatrix} -\beta_1 \\ -\beta_3 \\ -\beta_5 \\ -\beta_7 \end{pmatrix} = \begin{pmatrix} -\beta_1 \\ -\beta_3 \\ -\beta_1\beta_6 - \beta_5 \\ -\beta_3\beta_8 - \beta_7 \end{pmatrix} \tag{8}$$

As $\beta_6 = \beta_8$, we have a system of 4 equations in 5 unknowns:

(i) $\pi_{11} = -\beta_1$

(ii) $\pi_{21} = -\beta_3$

(iii) $\pi_{31} = -\beta_1\beta - \beta_5$

(iv) $\pi_{41} = -\beta_3\beta - \beta_7$

Without further restrictions the system is not identified. However, if time effects are ruled out, we can also assume

(v) $\beta_5 = \beta_7 = \beta_0$,

and the system is identified if $\beta_1 \neq \beta_3$.[9] If these conditions are met, availability of exogenous variables z is not necessary for the identification of β.

This condition can be recovered by exploiting the rank condition in Wegge (1965) and Kelly (1975). It is possible to show that the cross-equation constraints in the equations of interest allow identification of the parameter of interest β in the *full* system (7), if further assumptions are made either on the time evolution of y or on the absence of homoschedasticity.[10]

In the next Section, we set forth the identifiability conditions in the general case. The Monte Carlo experiment reported in Section 3 will allow us to compare the performance of Zellner-type estimation of the system (7) with the case when the external instruments are available, as well as with respect to traditional estimation methods.[11]

[9] See Appendix A for details.

[10] As for estimation, the argument by Lahiri and Schmidt (1978) guarantees consistency of an iterated SURE, because the right hand side variable in the first equation is exogenous (the first equation only contains a constant term on the right hand side) and the model is linear. However, as the system is misspecified (z are omitted), the estimates are not equivalent to the ML estimation of the full system (i.e. the system that also exploits the information in the variable z). Consistency is nonetheless achieved as, by assumption, the omitted variable (z) is truly exogenous.

[11] From a practical point of view, the model is easily estimated by standard econometric software by iterating to convergence a FGLS estimator of the reduced system, imposing proper cross-equation constraints. Of course, convergence is not guaranteed. Line search maximization of the likelihood, analogous to Dagenais (1978) and Calzolari *et al.* (1987) would be helpful in order to improve convergence.

2.1 Identifiability criteria in a general context

The works of Wegge (1965) and Kelly (1975) set forth a necessary and sufficient condition for the identification of a simultaneous system of equations. Let us consider the general case of a system of M equations involving K variables. Adopting Wegge (1965)'s formulation and notation, the system can be written as $u_i = Aw_i$, with u_i a M-component column vector of structural disturbances with variance covariance matrix Σ, $A = \{a_{ij}\}$, $i = 1, .., M$, $j = 1, ..., K$, and w_i a K-component vector of variables (both endogenous and exogenous variables are included in w_i). As the total number of variables in the model needs to be greater than the number of endogenous variables, we require $K > M$.

Wegge (1965) considers the general case in which the model parameters (both the components of the matrix A or the components of the variance covariance matrix of the error terms Σ) are subject to a priori restrictions (i.e. continuous differentiable functions). Kelly (1975) specifies the general approach to the case of linear cross-equation constraints. In this section we apply the identifiability conditions proposed by Wegge (1965) and Kelly (1975) to the panel data context.[12]

Let us consider the matrix $P = [A\ \Sigma]$. The vector of parameters can be built by considering the vectorization of matrix P, v_P (i.e. v_P is a row vector whose elements are taken row-wise from P):[13]

$$v_P = [a_{11}, ..., a_{1K}, \sigma_{11}, ..., \sigma_{1M}, a_{21}, ..., a_{2K}, \sigma_{21}, ..., \sigma_{2M}, ..., a_{M1}, ..., a_{MK}, \sigma_{M1}, ..., \sigma_{MM}]$$

The a priori restrictions (normalizing conditions, exclusion restrictions, and cross-equation constraints) can be written as

$$\varphi_r(v_P) = 0 \quad \text{with} \quad r = 1, ..., R \tag{9}$$

The true set of parameters v_P^0 satisfies the system: $\varphi_r(v_P^0) = 0$ $(r = 1, ..., R)$.

Let us consider a non-singular matrix F, and the transformed system $FAw_i = FA$. The transformed system needs to satisfy the a priori restrictions, i.e. $\varphi_r(v_{FP}) = 0$, in which $FP = [FA\ F\Sigma F']$. The system is identified if the only admissible matrix is $F = I$.

Wegge (1965) provides a solution based upon the rank of the $R \times M^2$ Jacobian matrix

$$J(F) = \left[\frac{\partial \varphi_r(FP)}{\partial v_F}\right] \tag{10}$$

in which v_F denotes the row-vectorization of matrix F.

If the matrix $J(I)$ has rank M^2, then the system is identified.[14]

In the case of panel data with T available time periods, the 'restricted' system (7) can be written as:

$$\begin{cases} x_{i1} &= \beta_1 + \tilde{u}_{i1} \\ x_{i2} &= \beta_3 + \tilde{u}_{i2} \\ \vdots & \\ x_{iT} &= \beta_{2T-1} + \tilde{u}_{iT} \\ y_{i1} &= \beta_{2T+1} + x_{i1}\beta_{2T+2} + u_{i,T+1} \\ \vdots & \\ y_{iT} &= \beta_{2T+2T-1} + x_{iT}\beta_{2T+2T} + u_{i,2T} \end{cases} \tag{11}$$

with $\tilde{u}_{it} = z_{it}\beta_{2t} + u_{it}$ $(t = 1, ..., T)$, $M = 2T$ endogenous variables, and $K = 2T + 1$ variables in the system.

[12] Please refer to Wegge (1965) and Kelly (1975) for details and proofs.

[13] The variance covariance matrix Σ is assumed to be symmetric, i.e. $\sigma_{ij} = \sigma_{ji}$. Further, restrictions are assumed to be symmetric in the arguments σ_{ij} and σ_{ji} (see Wegge, 1965 for details).

[14] Therefore the number of restrictions R needs to be at least equal to M^2.

For the general case of T observations, the following constraints can be imposed on the matrix of coefficients:

(a) $2 \times T$ normalization constraints ($a_{ii} - 1 = 0$, $i = 1, ..., M$), one for each endogenous variable in the system;

(b) $[T + (T-1)] \times T + [(T-1) + (T-1)] \times T$ exclusion restrictions, as the T equations that characterize x exclude all the y (T) and the x in the other time periods ($T-1$); and the T equations that characterize y exclude all the y and x in the other time periods ($T-1$ in both cases);

(c) $(T-1)$ cross-equation constraints as $\beta_{2T+2\tau} = \beta$ for each $\tau = 1, ..., T$ (the coefficient of the variable x is constant over time).[15]

Overall, the number of constraints (a) + (b) + (c) sums to $4T^2 - 1$, whereas at least $4T^2$ constraints would be needed to achieve identification.[16]
Therefore, we need to impose at least one more constraint. Two options are viable.
First, if time evolution independent from x can be excluded in y, $T-1$ cross-equation restrictions arise in system (11), i.e. we also have:[17]

(d) $(T-1)$ cross-equation constraints on $\beta_{2T+2\tau-1}$, $\tau = 1, ..., T$: $\beta_{2T+1} = \beta_{2T+3} = ... = \beta_{2T+2T-1}$.

therefore allowing identification.
As a drawback, the rank of the matrix $J(I)$, built exploiting the constraints (a) to (d), depends on the value of the parameters β_{2t-1}, $t = 1, ..., T$.[18]
In particular, if the condition $\beta_{2T+2t-1} = \beta_{2T+2s-1}$ ($t \neq s$) is imposed on the system parameters,[19] then the condition $\beta_{2t-1} \neq \beta_{2s-1}$ is needed for system identification.
As an alternative option, the structure of the variance covariance matrix can be exploited. Under an assumption of homoschedasticity in α_i and e_{it}, the following $T-1$ conditions hold:

(d') $(T-1)$ cross-equation constraints on the variance of the error term of the equation describing y: $V(u_{i,T+1}) = V(u_{i,T+2}) = ... = V(u_{i,T+T})$.

In this case, the rank of the matrix $J(I)$ depends on the relationship between the error terms in $\tilde{u}_{is}$ and in $u_{i,T+t}$ at different points in time. Particularly, if the condition $V(u_{i,T+t}) = V(u_{i,T+s})$ ($t \neq s$) is imposed, then $E(\tilde{u}_{is} u_{i,T+s}) - E(\tilde{u}_{it} u_{i,T+t}) \neq 0$ is needed for identification.

[15] Let us denote k the number of endogenous regressors included in the model among the explanatory variables (i.e. excluding y). In the case of $k > 1$ endogenous variables we have: (a) $(k+1)T$ normalization constraints, (b) $kT[T + (k-1)T + (T-1)] + T[k(T-1) + (T-1)]$ exclusion restrictions; (c) $k(T-1)$ cross-equation restrictions on β, for a total of $(k+1)^2T^2 - k$ restrictions in which at least $(k+1)^2T^2$ would be needed to achieve identification.

[16] The number of exclusion restrictions available is different if a dynamic framework is considered. Particularly, when the AR(1) specification is considered, i.e. $y_{it} = \gamma y_{it-1} + x'_{it}\beta + \varepsilon_{it}$, the number of restrictions is equal to the number of restrictions we count in the static framework minus $T-1$ exclusion restrictions due to the presence of y_{it-1} in the equations ($t = 2, ...T$). $T-2$ additional restrictions are available as the autoregression coefficient γ is constant over time. As a result, when x_{it} is a scalar, the sum of exclusion restrictions in a dynamic framework is $4T^2 - 2$.

[17] Therefore if $k > 1$, we need $k \leq T-1$ to achieve consistency, as we need the number of constraints to be at least $(k+1)^2T^2$: $(k+1)^2T^2 - k + (T-1) \geq (k+1)^2T^2$.

[18] As an example, when $T = 2$ in order to achieve identification of the system we need to impose that the intercepts of the equation linking y to x does not change over time. The matrix $J(I)$ is of full rank if $\beta_1 - \beta_3 \neq 0$, i.e. we are imposing time variation in x, whereas time variation (independent of x) is not allowed in y. See Appendix A for details.

[19] Recall that $\beta_{2T+2t-1}$ represent the intercept of the equation that relates y_{it} to x_{it}.

3 Estimation performance: Monte Carlo experiments

In this Section we perform a set of Monte Carlo experiments in order to show that, in cases when the conditions set forth in Section 2 are satisfied, Zellner-type estimation of the system of equations (11) can also provide performance gains in finite sample with respect to the most widely adopted estimation techniques.

We will take into consideration a linear static model in which both the strict exogeneity assumption and the orthogonality condition are violated, as the regressor is contemporaneously related to both the individual effect *and* the idiosyncratic component, and it exhibits a trend over time. Estimation is performed using fixed effect (FE) estimator (expected to be biased), GMM estimators that properly account for the endogeneity issue, along with a Zellner-type estimator of the system of equations (11), labeled SIM-EQ. Full Maximum Likelihood (FIML) estimator is reported as a benchmark (infeasible, as it considers full observability of the instruments z).

The dependent variable is defined as in equation (1) with $\varepsilon_{it} = \alpha_i + e_{it}$ with α_i and e_{it} distributed as a $N(0,1)$ independent of each other. The data generating process is set as follows:

$$x_{it} = \gamma_1 + \gamma_2 z_{it} + \alpha_i + e_{it} + w_{it} \quad (12)$$

$$y_{it} = \beta_1 + \beta_2 x_{it} + \alpha_i + e_{it}, \quad i = 1, ..., N; t = 1, ..., T \quad (13)$$

where $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 1$, z_{it} is a strictly exogenous variable exhibiting a trend over time ($z_{it} = 1 + t + \omega_{it}$), and w_{it} and ω_{it} are independent and normally distributed with mean zero and unit variance. As a result x_{it} is contemporaneously related to α_i and e_{it}.

When z are not observable, the FE estimator will provide biased estimates of the regression coefficient β_2, whereas it is possible to solve the endogeneity bias by applying a GMM framework. Results of Monte Carlo experiments are reported in Table 1. The estimation method based on the proposed identification strategy is labeled 'SIM-EQ', whereas 'GMM-AB' and 'GMM-BB' denote respectively the GMM approach proposed by Arellano and Bond (1991) and Blundell and Bond (1998).[20]

The Table also reports the Monte Carlo mean and standard error of the FIML estimator that considers the *full* structure of the system of equations, i.e. it also exploits the information about z_{it} (therefore, it is an "infeasible FIML"). This is the efficient estimator when information about z_{it} is available (and the error terms are normally distributed). Unfortunately, this is unlikely to be the case with real data, in which the availability of external instruments is usually severely limited, and the availability of repeated observations on the same unit is a source of "internal" instruments. Still, it provides an useful benchmark to compare the performance of the different estimators.

		FE	GMM-AB	GMM-BB	SIM-EQ	FIML
$T = 3$	Mean	1.249	.9974	1.003	.9998	.9998
	Std.Dev.	.0094	.0534	.0479	.0208	.0146
$T = 5$	Mean	1.182	1.001	1.002	.9999	.9999
	Std.Dev.	.0059	.0168	.0162	.0101	.0086
$T = 10$	Mean	1.082	1.000	1.001	1.000	1.000
	Std.Dev.	.0028	.0045	.0044	.0035	.0034

Table 1: Mean and standard error of Monte Carlo experiments: estimates of β_2 (true value equal to 1), $N = 1,000$.

[20]See also Arellano and Bover (1995). As error terms are homoschedastic, first stage estimates are considered.

		FE	GMM-AB	GMM-BB	SIM-EQ	FIML
$T = 3$	Mean	1.126	.9986	.9986	.9999	.9998
$\rho = 0.5$	Std.Dev.	.0080	.0307	.0199	.0108	.0095
$T = 5$	Mean	1.067	1.000	1.001	.9998	.9998
$\rho = 0.5$	Std.Dev.	.0041	.0089	.0067	.0049	.0047
$T = 10$	Mean	1.023	1.000	1.000	1.000	1.000
$\rho = 0.5$	Std.Dev.	.0017	.0024	.0022	.0017	.0017
$T = 3$	Mean	1.047	1.000	.9991	1.000	1.000
$\rho = 0.9$	Std.Dev.	.0061	.0172	.0101	.0068	.0062
$T = 5$	Mean	1.015	1.000	1.000	.9999	.9999
$\rho = 0.9$	Std.Dev.	.0025	.0042	.0030	.0026	.0025
$T = 10$	Mean	1.003	1.000	1.000	1.000	1.000
$\rho = 0.9$	Std.Dev.	.0007	.0008	.0008	.0007	.0007

Table 2: Mean and standard error of Monte Carlo experiments: estimates of β_2 (true value equal to 1), x follows an autoregressive structure, $N = 1,000$

As expected, the FE approach provides biased estimates, with the magnitude of the bias decreasing as T increases. When endogeneity is properly accounted for, the bias disappears.[21] However, the SIM-EQ estimator exhibits lower standard errors than GMM-AB and GMM-BB. In the case of $T = 10$, system estimator matches the performance of the "infeasible FIML" estimator that also employs additional information (i.e. the instruments z_{it}).
As a second experiment, we consider an autoregressive process for the generation of x_{it}:

$$x_{it} = \rho x_{it-1} + 1 + z_{it} + \alpha_i + e_{it} + w_{it} \tag{14}$$

The first observation is generated as: $x_{i0} = (1 + \alpha_i)/(1 - \rho) + (w_{i0} + e_{i0})/\sqrt{1 - \rho^2}$, and then discarded during estimation. Results of Monte Carlo experiments are reported in Table 2. We consider two different values of the autoregressive parameter: $\rho = 0.5, 0.9$. Results from the previous experiment are largely confirmed.
In this case the "infeasible FIML" estimator considers the relationship between x_{it} and x_{it-1} and z_{it}, as well as the equation of x_{i0}. In this framework, the performance of the SIM-EQ estimator is better the larger is the value of ρ, as the loss of information due to ignoring the presence of x_{it-1} as a determinant of x_{it} is smaller when the autocorrelation parameter is larger (the larger ρ, the larger the information contained in x_{it-1} that is 'passed' to x_{it}).

3.1 Dynamic case

Consider the dynamic model

$$y_{it} = \gamma y_{it-1} + \alpha_i + e_{it} = \gamma y_{it-1} + \varepsilon_{it} \tag{15}$$

in which we suppose that y_{it} is observed for $t = 0, 1, ..., T$ (i.e. $T + 1$ observations on y are available).

[21] When running GMM estimation, all available instruments are considered. Estimation is performed using own-written programs in Fortran 77. Also STATA has been used, and the command `xtabond2` is considered for GMM estimation (Roodman, 2006). The SIM-EQ strategy is easily implemented in STATA using the `sureg` command with suitable constraints on the model parameters.

Of course, if $\alpha_i = 0$ for each i and e_{it} is not correlated over time, the system of T equations defined by (15) is recursive and therefore identified (e.g., Lahiri and Schmidt 1978, switching the role of i and t).

If $\alpha_i \neq 0$, then we can expect y_{i0} to be endogenous due to correlation with α_i. Endogeneity of y_{i0} can also arise in cases when e_{it} is autocorrelated over time. In both cases, the variance-covariance matrix of e_i is full and identification of the system is not ensured (e.g., Greene 2003, ch. 15).

By exploiting the cross-equation constraints, coming from the time-invariant coefficient γ, it is possible to show that identification is achieved also in this case if the mean of y_{i0} is different from zero (see Appendix B for details).

The intuition in this case is to add an equation for the linear projection of the first observation

$$y_{i0} = \gamma_0 + \varepsilon_{i0} \tag{16}$$

If $\gamma_0 \neq 0$ the system is identified.[22] Observability of y_{i0} and y_{i1} is enough for identification of γ. In a more general case with intercept

$$y_{it} = \mu + \gamma y_{it-1} + \alpha_i + e_{it} = \mu + \gamma y_{it-1} + \varepsilon_{it} \tag{17}$$

with an endogenous y_{i0}, at least $T = 2$ (observability of y_{i0}, y_{i1} and y_{i2}) is needed for identification. Still, equation (16) is added to the system. In Appendix B we show that the system defined by (16) and (17) for $t = 1, ..., T$ is identified if $\gamma_0 \neq \mu/(1-\gamma)$.

[22]See also Calzolari and Magazzini (2011).

A Identification when T=2

The 'restricted' system is (see text – system 7):

$$\begin{cases} x_{i1} &= \beta_1 + \tilde{u}_{i1} \\ x_{i2} &= \beta_3 + \tilde{u}_{i2} \\ y_{i1} &= \beta_5 + x_{i1}\beta + u_{i3} \\ y_{i2} &= \beta_7 + x_{i2}\beta + u_{i4} \end{cases} \tag{18}$$

In Wegge (1965) and Kelly (1975) notation, the structural system of equations can be written as:

$$Aw_i = u_i \tag{19}$$

where $w_i = [x_{i1}, x_{i2}, y_{i1}, y_{i2}, 1]'$, $u_i = [\tilde{u}_{i1}, \tilde{u}_{i2}, u_{i1}, u_{i2}]'$, and

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{15} \\ a_{21} & a_{22} & \dots & a_{25} \\ \vdots & \vdots & \ddots & \vdots \\ a_{41} & a_{42} & \dots & a_{45} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & -\beta_1 \\ 0 & 1 & 0 & 0 & -\beta_3 \\ -\beta & 0 & 1 & 0 & -\beta_5 \\ 0 & -\beta & 0 & 1 & -\beta_7 \end{pmatrix} \tag{20}$$

In order to apply the identification conditions set forth in Wegge (1965) and Kelly (1975), we need at least M^2 conditions, where M is the number of endogenous variables in the system (in our case equal to 4: $x_{i1}, x_{i2}, y_{i1}, y_{i2}$). The available restrictions are: (a) 4 normalizing conditions: $a_{ii} - 1 = 0, i = 1, ..., 4$, corresponding to the ones in the matrix A, (b) 10 exclusion restrictions, corresponding to the zeros in the matrix A: $a_{12} = a_{13} = a_{14} = 0$, $a_{21} = a_{23} = a_{24} = 0$, $a_{32} = a_{34} = 0$, and $a_{41} = a_{43} = 0$; (c) one cross equation restriction spanning from standard panel data modeling in which the parameter β of the relationship between y and x is assumed to be constant over time: $a_{31} = a_{42}$ or $a_{31} - a_{42} = 0$. The total number of available restrictions is 15, whereas the rank condition requires at least 16 restrictions.
In order to identify the system one additional condition is needed, and we rule out trend in y (once x is accounted for): (d) $a_{35} = a_{45}$.
By building the Jacobian matrix of restrictions J as described in Kelly (1975),[23] that specifically consider the case of linear cross-equation constraints, a necessary and sufficient condition for identification (named Wegge's criterion in Kelly's paper), is that the rank of J equals M^2.

[23]Please refer to the original paper by Wegge (1965) and Kelly (1975) for details about the computation of the Jacobian matrix.

In our case, matrix J ($M^2 \times M^2 = 16 \times 16$) is:[24]

$$J(I) = \left(\begin{array}{cccc|cccc|cccc|cccc}
1 & 0 & -\beta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & -\beta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\hline
0 & 1 & 0 & -\beta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 1 & 0 & -\beta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -\beta & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -\beta & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -\beta & 0 & 0 & -1 & 0 & \beta \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\beta_1 & -\beta_3 & -\beta_0 & -\beta_0 & \beta_1 & \beta_3 & \beta_0 & \beta_0
\end{array}\right)$$

By applying the Sarrus rule, it is possible to show that the $\det(J) = \beta_1 - \beta_3$, therefore J is full rank (and the system parameters are identified) unless $\beta_1 = \beta_3$.
Let us consider the case when we allow a_{35} and a_{45} to differ and the additional condition to achieve identification is built on the basis of the variance covariance matrix of the error terms in the system.
Denote $\sigma_{jk} = E(u_j u_k')$, with u denoting irrespectively u and $\tilde{u}$. We write the condition $V(u_{i3}) = V(u_{i4})$ as $\sigma_{33} = \sigma_{44}$.
Only the last row of the matrix $J(I)$ is affected by the change in the restriction set and it is changed to:

$$J(I)_{[16,\cdot]} = \left(\begin{array}{ccc|ccc|cccc|cccc} 0 & \dots & 0 & 0 & \dots & 0 & 2\sigma_{13} & 2\sigma_{23} & 2\sigma_{33} & 2\sigma_{43} & -2\sigma_{14} & -2\sigma_{24} & -2\sigma_{34} & -2\sigma_{44} \end{array}\right)$$

The condition for full rank of $J(I)$ now becomes $\sigma_{24} - \sigma_{13} \neq 0$.

B Dynamic modeling

Let us consider the dynamic model

$$y_{it} = \gamma y_{it-1} + \alpha_i + e_{it} = \gamma y_{it-1} + \varepsilon_{it} \tag{21}$$

Let us start with $T = 1$, i.e. two observations are available on the dependent variable y_{i0} and y_{i1}.
As previous observations on y are not available we are not allowed to write the dynamic equation that characterizes y_{i0}, however we can consider the linear projection $y_{i0} = \gamma_0 + \varepsilon_{i0}$. We investigate the identification of the following system:

$$\begin{aligned} y_{i0} &= \gamma_0 + \varepsilon_{i0} \\ y_{i1} &= \gamma y_{i0} + \varepsilon_{i1} \end{aligned} \tag{22}$$

[24]The row ordering corresponds to one of constraints (a)-(d) listed in the text.

By using the notation in Wegge (1965) and Kelly (1975) we can write the system as $Aw_i = \varepsilon_i$ where

$$A = \begin{pmatrix} 1 & 0 & -\gamma_0 \\ -\gamma & 1 & 0 \end{pmatrix} \quad w_i = (y_{i0}, y_{i1}, 1)' \;\; \varepsilon_i = (\varepsilon_{i0}, \varepsilon_{i1})'$$

The matrix $J(I)$ introduced in Appendix A is:

$$J(I) = \left(\begin{array}{cc|cc} 1 & -\gamma & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & -\gamma_0 & 0 \end{array}\right)$$

As $\det(J(I)) = -\gamma_0$, the system (22) is identified if $\gamma_0 \neq 0$.
Now let us consider the case in which we allow an intercept in the model. In this case, if we are not willing to impose covariance restrictions, at least $T = 2$ is needed to achieve identification. We consider the system:

$$\begin{aligned} y_{i0} &= \gamma_0 + \varepsilon_{i0} \\ y_{i1} &= \mu + \gamma y_{i0} + \varepsilon_{i1} \\ y_{i2} &= \mu + \gamma y_{i1} + \varepsilon_{i2} \end{aligned} \tag{23}$$

We let $w_i = (y_{i0}, y_{i1}, y_{i2}, 1)$ and

$$A = \begin{pmatrix} 1 & 0 & 0 & -\gamma_0 \\ -\gamma & 1 & 0 & -\mu \\ 0 & -\gamma & 1 & -\mu \end{pmatrix}$$

We have:

$$J(I) = \left(\begin{array}{ccc|ccc|ccc} 1 & -\gamma & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -\gamma & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 1 & -\gamma & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\gamma & 0 \\ \hline 0 & 0 & 0 & 1 & -\gamma & 0 & 0 & -1 & \gamma \\ \hline 0 & 0 & 0 & -\gamma_0 & -\mu & -\mu & \gamma_0 & \mu & \mu \end{array}\right)$$

with $\det(J(I)) = \gamma_0(1-\gamma) - \mu$, that is, the system (23) is identified if $\gamma_0 \neq \mu/(1-\gamma)$.

References

Ahn, S., and P. Schmidt (1995): Efficient estimation of models for dynamic panel data, *Journal of Econometrics* **68**, 5–27.

Anderson, T.W. and Hsiao, C.: 1981, Estimation of dynamic models with error components, *Journal of the American Statistical Association* **76**, 598–606.

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

Arellano, M. (1990): Testing for Autocorrelation in Dynamic Random Effects Models, *The Review of Economic Studies* **57**(1), 127–134.

Arellano, M., and S. Bond (1991): Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economics and Statistics* **58**, 277–297.

Arellano, M., and O. Bover (1995): Another Look at the Instrumental-Variable Estimation of Error-Components Models, *Journal of Econometrics* **68**, 29–52.

Arellano, M., and B. Honoré (2001): Panel Data Models: Some Recent Development, in Heckman, J.J., and E. Leamer (eds.), *Handbook of econometrics* **5**, 3229–3296.

Berzeg, K. (1979): The Error Components Model. Conditions for the Existence of Maximum Likelihood Estimates, *Journal of Econometrics* **10**, 99–102.

Bhargava, A., and J.D. Sargan (1983): Estimating dynamic random effects models from panel data covering short time periods, *Econometrica* **51**, 1635–1659.

Blundell, R., and S. Bond (1998): Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics* **87**, 115–143.

Calzolari, G., L. Panattoni, and C. Weihs (1987): Computational Efficiency of FIML estimation, *Journal of Econometrics* **36**, 299–310.

Calzolari, G. and L. Magazzini (2011): Moment Conditions and Neglected Endogeneity in Panel Data Models, Department of Economics, University of Verona, Working Paper n. 2.

Dagenais, M.G. (1978): The Computation of FIML Estimates as a Iterative Generalized Least Squares Estimates in Linear and Nonlinear Simultaneous Equation Models, *Econometrica* **46**, 1352–1362.

Greene, W.: 2003, *Econometric analysis*, 6th ed., Prentice Hall.

Kelly, J.S. (1975): Linear Cross-Equation Constraints and the Identification Problem, *Econometrica* **43**(1), 125–140.

Lahiri, K. and Schmidth, P. (1978): Estimation of Triangular Structural Systems, *Econometrica* **46**(5), 1217–1221.

Mundlak, Y. (1978): On the Pooling of Cross Section and Time Series Data, *Econometrica* **46**, 69–86.

Roodman, D. (2006): How to Do xtabond2: An Introduction to Difference and System GMM in Stata, Working Paper Number 103, Center for Global Development.

Wegge, L.L. (1965): Identifiability Criteria for a System of Equations as a Whole, *Australian Journal of Statistics* 7, 67–77.