Diskussionsbeitrag aus dem
Fachbereich Wirtschaftswissenschaften
Universität Duisburg-Essen
Campus Essen

Nr. 171

Februar 2009

A Solution to the Problem of Too Many Instruments in Dynamic Panel Data GMM

Jens Mehrhoff

A Solution to the Problem of Too Many Instruments in Dynamic Panel Data GMM

Jens Mehrhoff[*], Deutsche Bundesbank

Abstract

The well-known problem of too many instruments in dynamic panel data GMM is dealt with in detail in Roodman (2009, Oxford Bull. Econ. Statist.). The present paper goes one step further by providing a solution to this problem: factorisation of the standard instrument set is shown to be a valid transformation for ensuring consistency of GMM. Monte Carlo simulations show that this new estimation technique outperforms other possible transformations by having a lower bias and RMSE as well as greater robustness of overidentifying restrictions. The researcher's choice of a particular transformation can be replaced by a data-driven statistical decision.

Keywords: Dynamic Panel Data, Generalised Method of Moments, Instrument Proliferation, Factor Analysis.

JEL: C13, C15, C23, C81.

[*]This paper represents the author's personal opinion and does not necessarily reflect the view of the Deutsche Bundesbank or its staff. Detailed results and descriptions of methodology are available on request from the author. Address for correspondence: Jens Mehrhoff, Statistics Department, Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main, Germany, Tel: +49 69 9566-3417, Fax: +49 69 9566-2941, E-mail: jens.mehrhoff@bundesbank.de, Homepage: www.bundesbank.de. The author would like to thank Professor Emeritus Dr. Volker Mammitzsch from the Faculty of Mathematics and Computer Science at the Philipps-Universität Marburg for valuable comments. All remaining errors are, of course, the author's sole responsibility.

1 The problem of too many instruments

Dynamic panel data (DPD) models have become increasingly popular in the last two decades. Nowadays the availability of micro level data, such as of firms or banks, enables researchers to identify economic relationships at a disaggregate level. Hence, the serious problem of aggregation bias (Lippi and Forni, 1990) can be avoided. However, the solution is not without a drawback: DPD bias. As Nickel (1981) has shown, the Least Squares Dummy Variables (LSDV) estimator has a non-vanishing bias for small T and large N. Anderson and Hsiao (1982) were the first to propose an unbiased DPD estimator with the notable trade-off between lag depth and sample size. It was not until Holtz-Eatkin et al. (1988) that an unbiased DPD estimator was constructed based on Generalised Method of Moments (GMM) (Hansen, 1982). The breakthrough came with Difference GMM by Arellano and Bond (1991), and System GMM by Arellano and Bover (1995) and Blundell and Bond (1998). In the meantime, Kiviet (1995) proposed a corrected LSDV estimator for balanced panels. However, one issue with regard to DPD GMM still remains unresolved; the number of instruments grows quadratically in T and GMM becomes inconsistent as the number of instruments diverges thus begging the question "what is the optimal set of instruments?"

Roodman (2009) addresses the problem of too many instruments. Increasing the sample size causes the number of instruments to proliferate as DPD GMM generates one instrument for each time period and lag available. Currently, there are two techniques in use to reduce the instrument count. One of them is limiting the lag depth, the other one is "collapsing" the instrument set. The former implies a selection of certain lags to be included in the instrument set, making the instrument count linear in T. The latter embodies a different belief about the orthogonality condition: it no longer needs to be valid for any one time period but still for each lag, again making the instrument count linear in T. A combination of both techniques makes the instrument count invariant to T. These transformations are deterministic ones of the instrument matrix, i.e. the transformation matrix consists of zeroes and ones. Besides the fact that no widely accepted rule of thumb for the instrument count exists, by choosing one of the aforementioned approaches, the researcher decides which transformation is to be used for the data.

The point in question is, “can we let the data decide how the transformation matrix should look?” The answer to this question is found by means of factor analysis of the instrument set and is shown to be “yes, we can.” The resulting DPD GMM estimator is characterised by both a lower bias and a lower root mean squared error (RMSE) than the standard techniques.

The remainder of the paper is organised as follows. Section 2 introduces the new estimation technique based on factorised instruments. Monte Carlo results for this estimator are presented in Section 3. The final section concludes.

2 A solution to this problem

Consider an autoregressive panel model of order one for the endogenous variable $y_{i,t}$, where α_i is a fixed effect and $\varepsilon_{i,t}$ is the error term.

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

The standard instrument set $\mathbf{Z}$ for the estimation of the autoregressive parameter β of Equation (1) with DPD GMM in first differences ($\Delta y_{i,t} = \beta \Delta y_{i,t-1} + \Delta \varepsilon_{i,t}$), which will be treated here exclusively without loss of generality but for simplicity of exposition, consists of lagged values of the endogenous variable, which are uncorrelated with the first differences of the error term.

$$\mathrm{E}(\mathbf{Z}'\boldsymbol{\Delta\varepsilon}) = \mathbf{0} \tag{2}$$

First, the conditions for consistency of the aforementioned techniques, along with a whole class of transformations, to reduce the instrument count are verified in the following theorem. Unlike other authors, who derive the limited or collapsed instrument set from first principles by considering interpretable orthogonality conditions, this paper applies transformation matrices to the standard instrument set which yield the desired results (cf. Appendix B). Proofs for this and the following theorem are to be found in Appendix A.

Theorem 1. *Let Equation (2) be valid. Then* $\mathrm{E}(\mathbf{Z}^{*\prime}\boldsymbol{\Delta\varepsilon}) = \mathbf{0}$ *with* $\mathbf{Z}^* = \mathbf{ZF}$ *for any deterministic transformation matrix* $\mathbf{F}$.

It follows from Theorem 1 that limiting the lag depth, collapsing the instrument set or both are valid transformations for consistent estimation of the parameter of interest. Moreover, any transformation, no matter if it lacks a sensible interpretation, satisfies the conditions of the theorem as long as it is deterministic.

Second, the aim of this paper is to introduce a new technique rather than to evaluate standards already in use. Hence, the focus here lies on stochastic transformations instead of deterministic ones. In order to solve the problem of instrument proliferation, this paper suggests the application of factor analysis – more precisely for the case in hand – principal components analysis (PCA) to the instrument set. PCA extracts the largest eigenvalues of the estimated covariance matrix of $\mathbf{Z}$ and assembles the corresponding eigenvectors in the matrix of component loadings $\mathbf{F}^*$, the transformation matrix. In this case, the transformation matrix is stochastic and Theorem 1 is no longer applicable. However, Theorem 2 provides a solution.

Theorem 2. *Let* $\mathbf{Z}$ *and* $\boldsymbol{\varepsilon}$ *be independent random variables. Then* $\mathrm{E}(\mathbf{Z}^{**\prime}\boldsymbol{\Delta\varepsilon}) = \mathbf{0}$ *with* $\mathbf{Z}^{**} = \mathbf{Z}\mathbf{F}^*$ *for any well-behaved stochastic transformation matrix* $\mathbf{F}^*$.

Here, well-behaved means Borel measurable. Theorem 2 is both more general and more specific than Theorem 1. The fact that it also holds true for deterministic $\mathbf{F}^* = \mathbf{F}$ makes it more general. It is more specific in the sense that it requires independence of $\mathbf{Z}$ and $\boldsymbol{\varepsilon}$ which is a stronger property than uncorrelatedness. This assumption is not too strong if the error term is thought of as being an exogenous shock.

3 Performance of factorised instruments

Judson and Owen (1999) provide Monte Carlo evidence that GMM is superior to other estimation techniques when it comes to DPD. Among others, their findings are: OLS produces biased estimates even for large T, the bias of LSDV decreases with T but may still be up to 20% of the true value even when $T = 30$, and also that the LSDV bias increases with the true value of the autoregressive parameter. Additionally, OLS is upward biased while LSDV is downward biased. Windmeijer (2005) adds to this list that GMM becomes more efficient when the lag depth is limited, and thus fewer instruments are employed in the estimation.

Table 1 presents biases and RMSEs from a Monte Carlo simulation of a one-step estimation of Equation (1) with parameter values of β in the range of zero to one. $\varepsilon_{i,t}$ is assumed to be standard normal, as is α_i. N is fixed at 100, T is 10, 20 and 30, respectively. The pre-sample period length is 30. The standard instrument set is either taken as it is, limited, collapsed or both, and additionally PCA has been applied to all four variants. The experiment is repeated 1,000 times.

The results confirm the findings of Judson and Owen (1999) and Windmeijer (2005). In addition, factorised instruments outperform all other techniques by having both a lower bias and RMSE, however, there are a few exceptions when $T = 10$. In general, factorisation of the limited and collapsed instrument set results in the lowest bias, while factorisation of the collapsed but unlimited instrument set yields the lowest RMSE. Biases are zero to the second decimal place or in relative terms less than 1%, RMSEs are zero to the first decimal place. The advantage of factorised instruments over standard ones is the condensation of the informational content of the instrument set into a much lower number of instruments employed in the estimation thus lowering the risk of overfitting endogenous variables but retaining almost all information. The next best approach is standard GMM with the instrument set being both limited and collapsed. Acceptable results can also be derived from a collapsed but unlimited instrument set in standard GMM. Limiting the lag depth on the one hand is a good idea as even if the autoregressive parameter is high, serial correlation will be low after a few periods and deeper lags are weak instruments, adding almost no new information for estimation. Collapsing the instrument set on the other hand also condenses the information in the instrument set into a lower number of instruments. The techniques most frequently used in applied DPD research, the untransformed instrument set and the limited one in standard GMM, are the worst choices, that is apart from the factorised variants of them. Both techniques are significantly downward biased, although the estimate still has the correct sign. Performance of their factorised variants is unacceptable; not even the correct sign can be expected. Explanations for the failure of the standard techniques can be found with recourse to the Sargan (1958) test of overidentifying restrictions (cf. Table 2). The failure of the factorised variants can be traced back to PCA and the Kaiser-Meyer-Olkin (Kaiser, 1970) measure of sampling adequacy (MSA) (cf. Table 3).

Table 1: Bias, standard error (SE) and RMSE for $\beta = .2$ and $\beta = .8$

		$T = 10$		$T = 20$		$T = 30$	
Method	Statistic	$\beta = .2$	$\beta = .8$	$\beta = .2$	$\beta = .8$	$\beta = .2$	$\beta = .8$
Least Squares							
OLS	Bias	$+.477$	$+.180$	$+.477$	$+.180$	$+.477$	$+.180$
	SE	.001	.000	.001	.000	.001	.000
	RMSE	.478	.180	.478	.180	.478	.180
LSDV	Bias	$-.136$	$-.243$	$-.064$	$-.111$	$-.042$	$-.070$
	SE	.001	.001	.001	.001	.001	.000
	RMSE	.140	.245	.068	.113	.045	.071
Standard GMM							
Untransformed	Bias	$-.080$	$-.539$	$-.146$	$-.624$	$-.199$	$-.681$
	SE	.002	.004	.001	.002	.001	.001
	RMSE	.101	.555	.151	.628	.201	.683
Limited (Ltd.)	Bias	$-.061$	$-.506$	$-.114$	$-.580$	$-.157$	$-.633$
	SE	.002	.005	.001	.002	.001	.002
	RMSE	.089	.528	.121	.585	.160	.635
Collapsed (Col.)	Bias	$-.014$	$-.373$	$-.017$	$-.296$	$-.017$	$-.257$
	SE	.002	.007	.001	.004	.001	.003
	RMSE	.070	.435	.047	.325	.039	.275
Ltd. & Col.	Bias	$-.001$	$-.172$	$-.007$	$-.159$	$-.007$	$-.137$
	SE	.002	.008	.001	.004	.001	.003
	RMSE	.071	.297	.044	.205	.036	.166
Factorised GMM							
Untransformed	Bias	$-.325$	$-.706$	$-.463$	$-.826$	$-.502$	$-.856$
	SE	.014	.018	.014	.015	.011	.013
	RMSE	.550	.913	.632	.945	.607	.949
Limited (Ltd.)	Bias	$-.165$	$-.534$	$-.300$	$-.646$	$-.399$	$-.760$
	SE	.008	.017	.010	.014	.010	.013
	RMSE	.305	.769	.447	.781	.501	.861
Collapsed (Col.)	Bias	$+.004$	$-.026$	$+.003$	$-.007$	$+.004$	.000
	SE	.002	.006	.001	.002	.001	.002
	RMSE	.059	.189	.035	.077	.029	.048
Ltd. & Col.	Bias	$+.002$	$+.005$	$+.002$	$-.002$	$+.003$	.000
	SE	.002	.007	.001	.003	.001	.002
	RMSE	.067	.217	.037	.084	.031	.055

Note: For the sake of brevity, results for values of the autoregressive parameter other than $\beta = .2$ and $\beta = .8$ are not displayed here. The results obtained for these values are similar to those presented above.

Table 2: Instrument count J and rejection frequency of the null hypothesis

	$T = 10$			$T = 20$			$T = 30$		
Method	J	$\beta = .2$	$\beta = .8$	J	$\beta = .2$	$\beta = .8$	J	$\beta = .2$	$\beta = .8$
Standard GMM									
Untransformed	36	.103	.202	171	.176	.400	406	.318	.605
Limited (Ltd.)	26	.096	.181	126	.140	.365	301	.228	.568
Collapsed (Col.)	8	.091	.166	18	.077	.169	28	.092	.185
Ltd. & Col.	4	.047	.097	9	.069	.096	14	.074	.099
Factorised GMM									
Untransformed	3	.080	.076	4	.064	.057	5	.070	.064
Limited (Ltd.)	3	.100	.109	4	.063	.064	5	.076	.072
Collapsed (Col.)	2	.000	.000	3	.000	.000	4	.000	.000
Ltd. & Col.	2	.000	.001	3	.000	.000	4	.000	.000

Table 2 shows the number of instruments employed in the estimation for each of the methods used and the proportions to which the validity of the overidentifying restrictions have been rejected at the nominal 5% significance level. It should be borne in mind that the power of the test is not weakened by many instruments. For limited instrument sets, the number of lags employed is set to be half of the available lags; for factorised instrument sets, the number of retained components has been fixed. Both choices are to a certain extent arbitrary.

Standard GMM with the untransformed or limited instrument set generates invalid overidentifying restrictions in an unacceptably high number of cases. This is due to the impossibility of fulfilling all restrictions simultaneously owing to the large number of instruments and the resulting overfitting of endogenous variables. Probabilities of rejection increase with β as well as with T. As it is known a priori that the null hypothesis of valid instruments or overidentifying restrictions is true in all cases, severe size distortions of the test become visible. While the test of the factorised variants of the collapsed (and limited) instrument set is undersized, rejecting the null hypothesis in virtually none of the cases, all tests of other instrument sets are oversized, some rather heavily.

Table 3 reports the explained variance and MSA from PCA. The explained variance states the proportion of the instrument set's variance that can be explained by the retained components. MSA is a statistical criterion to judge the

Table 3: Fraction of explained variance ρ and measure of sampling adequacy

		$T = 10$		$T = 20$		$T = 30$	
Method	Statistic	$\beta = .2$	$\beta = .8$	$\beta = .2$	$\beta = .8$	$\beta = .2$	$\beta = .8$
Untransformed	ρ	.398	.562	.247	.363	.200	.297
	MSA	.051	.859	.108	.930	.132	.948
Limited (Ltd.)	ρ	.350	.457	.197	.279	.154	.224
	MSA	.028	.776	.079	.901	.112	.931
Collapsed (Col.)	ρ	.700	.911	.670	.917	.669	.923
	MSA	.938	.999	.974	1.000	.981	1.000
Ltd. & Col.	ρ	.828	.968	.766	.966	.748	.967
	MSA	.926	.999	.977	1.000	.987	1.000

adequacy of the covariance matrix to be factorised; the closer it gets to one, the better. A value in the .90s is regarded as being "marvellous" in the literature.

The explained variance from PCA of the collapsed (and limited) instrument set is in the high .70s, low .80s for $\beta = .2$ and in the high .90s for $\beta = .8$. Almost all of the variation of the standard instrument set can be explained by much fewer components. Irrespective of β, PCAs of the untransformed or limited instrument set do not score appreciable values. This is the main reason why these procedures fail to result in plausible estimates (cf. Table 1). Although high MSAs can be achieved for $\beta = .8$, the explained variance remains low. MSAs for the first two procedures are close to one in all instances. The collapsed instrument set is much more suitable for PCA as each instrument is non-zero for all applicable observations, unlike untransformed instruments which are non-zero for just a single observation.

4 Directions for applied research

The Monte Carlo results strongly suggest the use of factorised instruments as these produce the lowest bias and RMSE. This generates an ultimate set of instruments and reduces the uncertainty researchers face in their choice of instruments. Furthermore, there is a clear recommendation to collapse the instrument set prior to factorisation or, if factorisation is not to be used at all, then at the very least the instrument set should be collapsed. To reiterate, this implies a deterministic transformation of the standard instrument set, and the factorised variant of this instrument set is the method of choice. Preferably, the lag depth is also limited. The lag limit should be chosen based on a priori information on the value of the autoregressive parameter, as serial correlation decreases exponentially. Most importantly, standard GMM suffers from instrument proliferation. The findings in this paper indicate that results of numerous applications of GMM in the literature may benefit from factorised instruments. LSDV should be applied only if the time dimension is much larger than 30, while pooled OLS should not be used at all in the estimation of DPD.

In applied research, the number of retained components from PCA can be derived from factor analytic criteria, such as MSA, and should be tested for their validity in the GMM framework. The methodology outlined here can be applied to System GMM or exogenous variables in a completely analogous fashion. It is reasonable to make use of the correlation between all instruments to lower the instrument count.

A Proof of theorems

Proof of Theorem 1. Using the definition of $\mathbf{Z}^*$ in Theorem 1 and Equation (2), the proposition follows directly from the linearity property of the expectation operator: $\mathrm{E}(\mathbf{Z}^{*\prime}\boldsymbol{\Delta\varepsilon}) = \mathrm{E}(\mathbf{F}'\mathbf{Z}'\boldsymbol{\Delta\varepsilon}) = \mathbf{F}'\mathrm{E}(\mathbf{Z}'\boldsymbol{\Delta\varepsilon}) = \mathbf{0}$. □

Proof of Theorem 2. Per definitionem of Theorem 2, $\mathbf{Z}$ and $\boldsymbol{\varepsilon}$ are a matrix and vector, respectively, of independent random variables, and thus Borel. For any pair $\phi(\cdot)$ and $\psi(\cdot)$ of Borel functions, this is also the case for $\phi(\mathbf{Z})$ and $\psi(\boldsymbol{\varepsilon})$.

$\widehat{\mathrm{Var}}(\mathbf{Z})$ is a positive semi-definite symmetric matrix meaning that all eigenvalues are real and non-negative. It is well-established that the sum and product of two real-valued measurable functions are measurable. That eigenvectors can be found in a Borel measurable fashion was shown by Azoff (1974, Corollary 4).

Hence, $\mathbf{Z}^{**} = \mathbf{Z}\Lambda(\mathbf{Z}) = \phi(\mathbf{Z})$, with $\mathbf{F}^* = \Lambda(\mathbf{Z})$ being the matrix of component loadings, and $\boldsymbol{\Delta\varepsilon} = \psi(\boldsymbol{\varepsilon})$ are independent random variables, too. Moreover, given quadratic integrability of $\mathbf{Z}^{**}$ and $\boldsymbol{\Delta\varepsilon}$, they are uncorrelated. The proposition follows from the fact that this can be the case if and only if $\mathrm{E}(\mathbf{Z}^{**\prime}\boldsymbol{\Delta\varepsilon}) = \mathbf{0}$ as $\mathrm{E}(\boldsymbol{\Delta\varepsilon}) = \mathbf{0}$. □

B Structure of transformation matrices

For the sake of exposition, let $T = 6$ and $i = 1, 2, \ldots, n$. Note that the first observation is dropped due to differencing.

Untransformed

The standard instrument set consists of lagged values of the endogenous variable; in particular, one instrument is generated for each time period and lag available. The instrument count is $J = (T-2)(T-1)/2 = 10$.

$$\mathbf{Z_i} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ y_{i,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & y_{i,2} & y_{i,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & y_{i,3} & y_{i,2} & y_{i,1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & y_{i,4} & y_{i,3} & y_{i,2} & y_{i,1} \end{bmatrix}$$

Limited (L)

Limiting the maximum lag depth of $y_{i,t-1}$ to $\tau = 2$, for example, gives as transformation matrix a block matrix of identity matrices up to dimension τ (for each time period, indicated by solid lines) separated by rows of zeroes (for excluded lags, indicated by dashed lines). Using this technique reduces the instrument count to $J^L = J - (T-2-\tau)(T-1-\tau)/2 = 7$.

$$\mathbf{Z_i^L} = \mathbf{Z_i}\mathbf{F^L} = \mathbf{Z_i} \left[\begin{array}{c|cc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ y_{i,1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & y_{i,2} & y_{i,1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & y_{i,3} & y_{i,2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & y_{i,4} & y_{i,3} \end{bmatrix}$$

Collapsed (C)

The transformation matrix for collapsing the instrument set is made up of identity matrices of increasing dimension stacked one upon the other (indicated by solid lines) with blocks of zero matrices to the right (indicated by dashed lines). By collapsing the instrument count is cut to $J^C = T - 2 = 4$.

$$\mathbf{Z_i^C} = \mathbf{Z_i F^C} = \mathbf{Z_i} \begin{bmatrix} 1 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ y_{i,1} & 0 & 0 & 0 \\ y_{i,2} & y_{i,1} & 0 & 0 \\ y_{i,3} & y_{i,2} & y_{i,1} & 0 \\ y_{i,4} & y_{i,3} & y_{i,2} & y_{i,1} \end{bmatrix}$$

Limited & Collapsed (LC)

When both techniques are combined, i.e. rows of zeroes from $\mathbf{F^L}$ and stacked identity matrices (now again only up to dimension τ) from $\mathbf{F^C}$, the instrument count becomes $J^{LC} = \tau = 2$.

$$\mathbf{Z_i^{LC}} = \mathbf{Z_i F^{LC}} = \mathbf{Z_i} \begin{bmatrix} 1 & 0 \\ \hline 1 & 0 \\ 0 & 1 \\ \hline 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ \hline 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ y_{i,1} & 0 \\ y_{i,2} & y_{i,1} \\ y_{i,3} & y_{i,2} \\ y_{i,4} & y_{i,3} \end{bmatrix}$$

References

Anderson, T.W., and Hsiao, C. (1982), "Formulation and estimation of dynamic models using panel data," *Journal of Econometrics*, 18, 47-82.

Arellano, M., and Bond, S. (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of Economic Studies*, 58, 277-297.

Arellano, M., and Bover, O. (1995), "Another look at the instrumental variable estimation of error-components models," *Journal of Econometrics*, 68, 29-51.

Azoff, E.A. (1974), "Borel measurability in linear algebra," *Proceedings of the American Mathematical Society*, 42, 346-350.

Blundell, R., and Bond, S. (1998), "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 87, 115-143.

Hansen, L.P. (1982), "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029-1054.

Holtz-Eatkin, D., Newey, W., and Rosen, H.S. (1988), "Estimating vector autoregressions with panel data," *Econometrica*, 56, 1371-1395.

Judson, R.A., and Owen, A.L. (1999), "Estimating dynamic panel data models: a guide for macroeconomists," *Economics Letters*, 65, 9-15.

Kaiser, H.F. (1970), "A second generation little jiffy," *Psychometrika*, 35, 401-415.

Kiviet, J.F. (1995), "On bias, inconsistency, and efficiency of various estimators in dynamic panel data models," *Journal of Econometrics*, 68, 53-78.

Lippi, M., and Forni, M. (1990), "On the dynamic specification of aggregated models," in: Barker, T., and Pesaran, M.H. (eds.), *Disaggregation in econometric modelling*, London: Routledge.

Nickel, S. (1981), "Biases in dynamic models with fixed effects," *Econometrica*, 49, 1417-1426.

Roodman, D. (2009), “A note on the theme of too many instruments,” *Oxford Bulletin of Economics and Statistics*, 71, 135-158.

Sargan, J.D. (1958), “The estimation of economic relationships using instrumental variables,” *Econometrica*, 26, 393-415.

Windmeijer, F. (2005), “A finite sample correction for the variance of linear efficient two-step GMM estimators,” *Journal of Econometrics*, 126, 25-51.