INSTRUMENT RELEVANCE AND EFFICIENT ESTIMATION WITH PANEL DATA

Rachid BOUMAHDI[*] and Alban THOMAS [†]

Abstract

This paper presents a straightforward efficiency-gain measure of instrumental-variable estimators for panel data regression models. Contrary to usual canonical correlation applied to the full set of endogenous regressors, we allow for measuring instrument relevance for separate endogenous regressors.

JEL Classification System: C23. J31.

Keywords: Panel data. Instrumental Variable. Efficiency.

---

1 Introduction

Cornwell and Rupert (1988) (CR) and Baltagi and Khanti-Akom (1990) (BKA) have investigated efficiency gains of instrumental variable estimators for panel data by applying the methods proposed by Hausman and Taylor (1981) (HT), Amemiya and MaCurdy (1986) (AM) and Breusch, Mizon and Schmidt (1989) (BMS). Fitting a wage equation on a Panel Study of Income Dynamics (PSID) data set, CR found that efficiency gains are limited to the coefficients of time-invariant endogenous variables. The authors claimed that (p.155), "The impact of the AM and BMS estimators falls primarily on time-invariant endogenous variables (like education) because the extra instruments employed by these methods are time-invariant." However, with same the data set, BKA found that efficiency gains are not limited to the coefficient on education, without providing an explanation for this result. BKA use the canonical correlation coefficient to compare different sets of instrumental variables, allowing one to calculate a correlation measure between instruments and the full set of endogenous regressors.

However, canonical correlations only measure instrument relevance for a set of endogenous regressors, and cannot be used to evaluate instrument relevance for a particular endogenous regressor. We generalize in this paper the measure of instrument relevance proposed by Shea (1997) and Godfrey (1999) to the case of panel data. This procedure measures relevance for each endogenous regressor separately, and we illustrate its use on the BKA empirical application.

2 A measure of instrument relevance

Consider the panel data regression model

$$y_{it} = X_{it}\beta + Z_i\lambda + \alpha_i + \varepsilon_{it}, \qquad i = 1, 2, ..., N; \ t = 1, 2, ..., T, \tag{1}$$

where the individual effect α_i is $IID(0, \sigma_\alpha^2)$, and ε_{it} is $IID(0, \sigma_\varepsilon^2)$ and is uncorrelated with (X_{it}, Z_i, α_i). β and λ are k and g vectors of coefficients. We assume T fixed and $N \to \infty$. Stacking all $T \times N$ observations we can write (1) as:

$$y = X\beta + Z\lambda + \alpha + \varepsilon. \tag{2}$$

If X and Z are uncorrelated with α, Generalized Least Squares (GLS) yields consistent and efficient parameter estimates. This estimator can be obtained by Ordinary Least Squares from:

$$Y = M\delta + \eta, \tag{3}$$

where $Y = \Omega^{-\frac{1}{2}}y$, Ω is the $NT \times NT$ variance-covariance matrix of the full equation residual $\alpha + \varepsilon$, $M = \left(\Omega^{-\frac{1}{2}}X : \Omega^{-\frac{1}{2}}Z\right)$, $\delta' = (\beta' : \lambda')$, $\eta = \Omega^{-\frac{1}{2}}(\alpha + \varepsilon)$. The GLS estimator is computed as:

$$\widehat{\delta}^{GLS} = \left(M'M\right)^{-1} M'Y, \tag{4}$$

Note that $\Omega^{-\frac{1}{2}}$ is proportional to $I_{NT} - (1-\theta)B$ where $B = I_N \otimes (1/T)V_T V_T'$ is the between operator, with $\theta^2 = \sigma_\varepsilon^2/(\sigma_\varepsilon^2 + T\sigma_\alpha^2)$, and V_T denotes a T vector of ones. The estimated variance-covariance matrix of the GLS estimator is:

$$V\left(\widehat{\delta}^{GLS}\right) = \widehat{\sigma}_{\eta_{GLS}}^2 \left(M'M\right)^{-1}, \tag{5}$$

where $\widehat{\sigma}_{\eta_{GLS}}^2 = \widehat{\eta}_{GLS}'\widehat{\eta}_{GLS}/(TN - k - g)$ and $\widehat{\eta}_{GLS}$ is the GLS residual. If we now assume that some variables in X or Z are correlated with α, consistent and efficient parameter estimates can still be obtained from an Instrumental Variable (IV) procedure. The general IV estimator is:

$$\widehat{\delta}^{IV} = \left(M'P_S M\right)^{-1} M'P_S Y, \tag{6}$$

where $P_S = S(S'S)^{-1}S'$ and S is the matrix of instruments. HT, AM and BMS procedures are based on a partition of $X = (X_1, X_2)$ and $Z = (Z_1, Z_2)$ and the assumption that α is uncorrelated with X_1 and Z_1. For instance, HT recommend an instrument matrix $S_{HT} = (WX_1 : WX_2 : BX_1 : Z_1)$, where $W = I_{NT} - B$ is the within operator. The covariance matrix estimate of $\widehat{\delta}^{IV}$ is:

$$V\left(\widehat{\delta}^{IV}\right) = \widehat{\sigma}_{\eta_{IV}}^2 \left(M'P_S M\right)^{-1}, \tag{7}$$

where $\widehat{\sigma}_{\eta_{IV}}^2 = \widehat{\eta}_{IV}'\widehat{\eta}_{IV}/(NT - k - g)$ and $\widehat{\eta}_{IV}$ is the IV residual.

Following Shea (1997) and Godfrey (1999), we consider estimation of a single parameter from δ by rewriting Equation (3) as:

3

$$Y = M_1 \delta_1 + M_2 \delta_2 + \eta, \tag{8}$$

where M_1 is $NT \times 1$ and M_2 is $NT \times (k + g - 1)$. Parameter δ_1 would for example correspond to the first variable in $\Omega^{-\frac{1}{2}} X_2$. Define $\underline{M}_1 = (I_{NT} - P_{M_2}) M_1$, $\overline{M}_1 = (I_{NT} - P_{\widehat{M}_2}) \widehat{M}_1$ and $\widehat{M}_j = P_{S_j} M_j$, $j = 1, 2$. These definitions imply that $\overline{M}'_1 \underline{M}_1 = \overline{M}'_1 \overline{M}_1$. Using the same idea as in Shea (1997) and Godfrey (1999) in the case of a linear multiple regression model, we can use as a measure of instrumental variable relevance, the population squared correlation between $\overline{M}_1$ and $\underline{M}_1$:

$$\rho_p^2 = \text{plim} \frac{\left(\overline{M}'_1 \underline{M}_1 \right)^2}{\left(\overline{M}'_1 \overline{M}_1 \right) \left(\underline{M}'_1 \underline{M}_1 \right)} = \text{plim} \frac{\overline{M}'_1 \overline{M}_1}{\underline{M}'_1 \underline{M}_1}. \tag{9}$$

In applied work, provided N tends to infinity, we can approximate ρ_p^2 by the following coefficient

$$R_p^2 = \frac{\overline{M}'_1 \overline{M}_1}{\underline{M}'_1 \underline{M}_1}, \tag{10}$$

which is directly related to the estimated parameter standard errors. To see this, consider the estimated variances $\widehat{\delta}_1^{GLS}$ and $\widehat{\delta}_1^{IV}$:

$$V\left(\widehat{\delta}_1^{GLS} \right) = \widehat{\sigma}_{\eta_{GLS}}^2 \left(\underline{M}'_1 \underline{M}_1 \right)^{-1} \quad \text{and} \quad V\left(\widehat{\delta}_1^{IV} \right) = \widehat{\sigma}_{\eta_{IV}}^2 \left(\overline{M}'_1 \overline{M}_1 \right)^{-1}. \tag{11}$$

Then, R_p^2 can be written as

$$R_p^2 = \frac{\widehat{\sigma}_{\eta_{IV}}^2 V\left(\widehat{\delta}_1^{GLS} \right)}{\widehat{\sigma}_{\eta_{GLS}}^2 V\left(\widehat{\delta}_1^{IV} \right)} = \frac{\overline{M}'_1 \overline{M}_1}{\underline{M}'_1 \underline{M}_1}. \tag{12}$$

Consequently, the measure of instrument relevance can be directly obtained by inspecting individual parameter (squared) standard errors. Note that in general, standard errors for GLS and IV are different, because they are computed using the actual regression residuals which will differ between GLS and IV, as these procedures yield different point coefficient estimates.

However, in the context of panel data, one can use $\widehat{\sigma}_\varepsilon^2 = \widehat{\varepsilon}'_W \widehat{\varepsilon}_W / (NT - k - g)$ where $\widehat{\varepsilon}_W$ is the within residual, as a consistent estimate for $\sigma_{\eta_{GLS}}^2$ and $\sigma_{\eta_{IV}}^2$. In this case, R_p^2 is simply the ratio of the GLS and IV variances of δ_1.

Naturally, if the IV and GLS variances errors are computed with regression residuals and are different, one cannot eliminate $\widehat{\sigma}^2_{\eta_{IV}}$ and $\widehat{\sigma}^2_{\eta_{GLS}}$, and in this case Equation (12) should be used.

3   Basic findings

Based on the same data set as in BKA, we compute our measure of instrument relevance R^2_p, and apply it to the BKA wage equation estimation results. The definition of variables used is[1]:

$LWAGE$: dependent variable, log of wage rate; OCC: blue-collar; IND: manufacturing; $SOUTH$: residence in the south; $SMSA$: residence in a SMSA; MS: married; EXP: years of full time work experience; WKS: weeks worked in the year; $UNION$: wage set by a union contract; ED: education level (years); FEM: female; BLK: Black.

In the first model specification of BKA (Table I, p.403) based on $X_1 = (WKS, SOUTH, SMSA, MS), X_2 = (EXP, EXP^2, OCC, IND, UNION), Z_1 = (FEM, BLK), Z_2 = (ED)$, standard errors of parameter estimates are the same for HT, AM and BMS, except for the coefficient on education. Similar results led CR to claim that IV efficiency gains are limited to the time-invariant regressor only. This claim can be confronted to Table 1 in our paper, where we report the population squared correlation coefficient between each endogenous variable and the instrument sets proposed by HT, AM and BMS[2]. The population squared correlation coefficient for all endogenous variables except ED is very stable from HT to AM and from AM to BMS. For education however, the coefficient is 4.57 percent, 8.52 percent and 11.27 percent for HT, AM and BMS respectively.

[INSERT TABLE 1 HERE]

---

[1]See Cornwell and Rupert (1988) for a description of these variables.

[2]Consistently with Equation (12), we use the same Within variance estimate $\widehat{\sigma}^2_\varepsilon$ in place of both $\widehat{\sigma}^2_{\eta_{GLS}}$ and $\widehat{\sigma}^2_{\eta_{IV}}$. Our results did not change significantly when we considered alternative estimates: $\widehat{\sigma}^2_\varepsilon = 0.0230$, $\widehat{\sigma}^2_{\eta_{GLS}} = 0.0228$, $\widehat{\sigma}^2_{\eta_{IV}} = 0.0231$ for HT and $\widehat{\sigma}^2_{\eta_{IV}} = 0.0230$ for AM.

Because this first specification of BKA is rejected by a Hausman test statistic based on the difference between the Within and the HT estimator, BKA proceeded by proposing another set of instruments, based on $X_1 = (OCC, SOUTH, SMSA, IND)$ and $Z_1 = (FEM, BLK)$ (see Table II of BKA, p. 404). For BMS, the estimator denoted BMS^a removes EXP, EXP^2 from the set of additional admissible instruments, while estimator BMS^b removes EXP, EXP^2 and $UNION$ from the set of additional instruments.[3]

The measure of instrument variable relevance corresponding to this specification is presented in Table 2. Here again, standard errors are very similar for time-varying variables. Efficiency gains seem to be limited to the time-invariant variable only, although these gains appear poorer than before (in the misspecified model). For the variable ED, the instrument relevance coefficient increases by 2.61 percent for AM relative to HT, by 1.88 percent for BMS^a relative to AM, and by 0.93 percent percent when we move from AM to BMS^b.

BKA use the canonical correlation coefficient for comparing the sets of instruments proposed by HT, AM and BMS. However, this coefficient only provides a global indication of the correlation between endogenous variables and instruments in the model. When employed by BKA, it indicates that BMS instruments are more correlated with endogenous variables than are the HT and AM instruments. However, the canonical correlation does not help to explain why the efficiency gains are concentrated more in some variables than in others.

[INSERT TABLE 2 HERE]

4 Concluding remarks

We have presented in this paper a method for assessing the efficiency gain associated with instrumental-variable procedures for panel data regression models. Contrary to BKA, it allows for the computation of efficiency gains on an individual parameter basis. This procedure is very easy to compute, as it uses only standard information from regression results, i.e., standard errors of selected parameter estimates.

---

[3]Complete results for all models in BKA are available from authors.

References

Amemiya, T. and T.E. MaCurdy, 1986. Instrumental–Variable Estimation of an Error–Component Model. *Econometrica* 54, 869-880.

Baltagi, B.H. and S. Khanti-Akom, 1990, On efficient estimation with panel data: an empirical comparison of instrumental variables estimators, Journal of Applied Econometrics 5, 401-406.

Breusch, T.S., G.E. Mizon and P. Schmidt, 1989, Efficient estimation using panel data, Econometrica 57, 695-700.

Cornwell, C. and P. Rupert, 1988, Efficient estimation with panel data: an empirical comparison of instrumental variables estimators, Journal of Applied Econometrics 3, 149-155.

Godfrey, L.G., 1999, Instrument relevance in multivariate linear models, Review of Economics and Statistics 81, 550-552.

Hausman, J.A and W.E. Taylor, 1981, Panel data and unobservable individual effects, Econometrica 49, 1377-1398.

Shea, J., 1997, Instrument relevance in multivariate linear models: a simple measure, Review of Economics and Statistics 79, 348-352.

Table 1: Estimated standard errors and instrument relevance coefficient (R_p^2) - Model 1

	GLS	HT	AM	BMS
EXP	*0.0024*	*0.0025*	*0.0025*	*0.0025*
		0.9216	**0.9216**	**0.9216**
EXP^2	*0.00005*	*0.00005*	*0.00005*	*0.00005*
		0.9999	**0.9999**	**0.9999**
OCC	*0.0136*	*0.0138*	*0.0138*	*0.0138*
		0.9712	**0.9712**	**0.9712**
IND	*0.0151*	*0.0154*	*0.0154*	*0.0154*
		0.9614	**0.9614**	**0.9614**
$UNION$	*0.0146*	*0.0149*	*0.0149*	*0.0149*
		0.9601	**0.9601**	**0.9601**
ED	*0.0141*	*0.0659*	*0.0483*	*0.0420*
		0.0457	**0.0852**	**0.1127**
Hausman test		$\chi^2(3)=14.5$	$\chi^2(13)=16.9$	$\chi^2(13)=13$
(p-value)		(0.002)	(0.2039)	(0.4478)

Standard errors are in italic, and instrument relevance ratios R_p^2 in bold face.

Table 2: Estimated standard errors and instrument relevance coefficient (R_p^2) - Model 2

	GLS	HT	AM	BMS^a	BMS^b
WKS	*0.0006*	*0.0006*	*0.0006*	*0.0006*	*0.0006*
		0.9999	**0.9999**	**0.9999**	**0.9999**
MS	*0.0187*	*0.0190*	*0.0190*	*0.0190*	*0.0190*
		0.9686	**0.9686**	**0.9686**	**0.9686**
ED	*0.0141*	*0.0212*	*0.0206*	*0.0202*	*0.0204*
		0.4423	**0.4684**	**0.4872**	**0.4777**
Hausman test		$\chi^2(3)=5.25$	$\chi^2(13)=14.74$	$\chi^2(13)=9.59$	$\chi^2(13)=7.56$
(p-value)		(0.1543)	(0.3238)	(0.7270)	(0.8710)

Standard errors are in italic, and instrument relevance ratios R_p^2 in bold face. See the text for the definition of BMS^a and BMS^b estimators. Results for $EXP, EXP^2, UNION$ are not reported here, as they are similar to those in Table 1.