Department of Economics
University of Southampton
Southampton SO17 1BJ
UK

Discussion Papers in Economics and Econometrics

ON THE ESTIMATION OF COVARIANCE MATRICES USING PANEL DATA ARTIFICIAL REGRESSIONS

Eleonora Patacchini

No. 0303

This paper is available on our website
http://www/soton.ac.uk/~econweb/dp/dp03.html

On the estimation of covariance matrices using panel data artificial regressions

Eleonora Patacchini
University of Southampton

Abstract

The use of artificial regressions to compute the variance of the difference of pairs of panel data estimators that cannot be ranked in terms of efficiency is considered. It is illustrated how it is possible to get (asymtotically) valid estimators of covariance matrices for differences between estimators when the assumption that the error term in the auxiliary model is IID is violated. We distinguish two possible deviations, one leading only to a non-spherical-within groups covariance matrix and the second leading to a non-spherical-between-groups covariance matrix also. It is shown to what extent the use of an artificial regression with panel data can lead to a robust estimator of the covariance matrix in the first case whereas it leads to a non valid estimator in the second. An alternative step by step procedure is presented.

Keywords: artificial regression models, panel data, covariance matrices estimates, hypothesis testing.

JEL Classification: C12, C13, C23

I am very grateful to Raymond O'Brien for comments and suggestions.

1 Introduction

One recurrent problem in applied studies is to compute the variance of the difference between two estimators that cannot be ranked in terms of efficiency. Facing this problem should be the rule rather than the exception in empirical studies because the assumptions underlying the application of standard results (e.g. Lemma 2.1 in Hausman, 1978) are too strong in many cases of study. For instance, the assumption of spherical disturbances is restrictive. To neglect this problem may lead to unreliable inference. However, theoretical solutions that require a knowledge of an advanced programming language (e.g. Lee, 1996) are often not suitable for the applied econometrician. The use of an artificial regression (e.g. MacKinnon, 1992) may be helpful in some cases but covariance matrices calculated using artificial regressions may not be asymptotically valid when the assumption that the error term in the auxiliary model is IID is violated (Davidson and MacKinnon, 1998).

This paper considers the use of artificial regressions to compute the variance of the difference between pairs of panel data estimators. It illustrates how it is possible to obtain valid estimates of covariance matrices between estimators when the assumption that the error term in the auxiliary model is IID is violated. These results can be used to perform Hausman tests (Hausman, 1978) robust to deviations from the classical errors assumption. The appealing feature of these methods for applied work is that they can be implemented in standard statistical packages.

Dealing with panel data, we distinguish two possible deviations from the assumption of spherical disturbances in the auxiliary model, one leading only to a non-spherical-within-groups covariance matrix and the second leading to a non-spherica-between-groups covariance matrix also. In other words, in the first case we still deal with a block-diagonal matrix whereas in the latter also non-zero elementss outside the diagonal blocks are allowed. It is shown to what extent the use of an artificial regression with panel data can lead to a robust estimator of the covariance matrix in the first case whereas it leads to a non valid estimator in the second. In those cases, an alternative step by step procedure is presented.

2 Within groups non spherical disturbances

O'Brien and Patacchini (2003) show that the use of a panel data artificial regression and the application of the White's formulae (White, 1984) can lead to consistent covariance matrix estimators that are robust to heteroscedasticity and/or autocorrelation within groups. They use these results to construct a robust formulation of the Hausman test for correlated effects based on the comparison between the Within Groups (WG) and the Between Groups (BG) estimators, and robust to deviations from the assumption of spherical disturbances.

However a similar procedure can be used for the comparison of a variety of panel data estimators. The condition to be satisfied in order to obtain a valid (robust) estimator of the covariance matrix is that the two estimators involved in the procedure have to be constructed using orthogonal transformations of the data. This ensures that the block diagonal structure of the covariance matrix is maintained. For instance, let us consider the comparison between the Generalized Instrumental Variables estimator using data in deviations from individual time-means, hereafter IVD estimator and the same Generalized Instrumental Variables estimator on the model in levels, hereafter IVL estimator. As in the comparison between the BG and the WG estimators, we deal with two different estimators that are obtained applying the same estimation method on data transformed in different ways. If we choose an IV estimator on the model in averages (*between groups* transformation) as IV estimator for the model in levels we deal again with two orthogonal transformations of the data. The use of an artificial regression will typically lead to the desired outcome.

Consider the general panel data model for individual i

$$\underset{(T\times 1)}{y_i} = \underset{(T\times K)}{X_i}\,\beta + \underset{(T\times 1)}{v_i}, \qquad i = 1, ..., N, \tag{1}$$

where the variance-covariance matrix of the v_i is

$$\Omega = \begin{pmatrix} \sigma_\eta^2 + \sigma^2 & \dots & \sigma_\eta^2 \\ \vdots & \ddots & \vdots \\ \sigma_\eta^2 & \dots & \sigma_\eta^2 + \sigma^2 \end{pmatrix} = \sigma^2 I_T + \sigma_\eta^2\ \iota\iota^{'},$$

and ι is a column vector of T ones. The $NT \times 1$ vector of disturbances has variance covariance matrix

$$\underset{(NT\times NT)}{\Sigma} = I_N \otimes \Omega.$$

This system of T equations in levels can be transformed into $(T-1)$ equations in deviations and one in averages. We obtain

$$\begin{cases} y_i^* = x_i^*\beta + v_i^* \\ \overline{y_i} = \overline{x_i}\beta + \overline{v_i}. \end{cases}$$

Estimating by IV the first group of equations, i.e. the ones in deviations from individual time means, we obtain the IVD estimator, i.e. $\widehat{\beta}_{ivd}$. Estimating by IV the average equation we obtain the IVL estimator, i.e. $\widehat{\beta}_{ivl}$.

Let

$$\beta_{ivd} = E\left(\widehat{\beta}_{ivd}\right)$$

and

$$\beta_{ivl} = E\left(\widehat{\beta}_{ivl}\right).$$

Rewrite the system as

$$\begin{cases} y_i^* = x_i^* \beta_{ivd} + \mu_i^* - x_i^* \beta_{ivl} + x_i^* \beta_{ivl} \\ \overline{y}_i = \overline{x}_i \widehat{\beta}_{ivl} + \overline{\mu}_i. \end{cases}$$

Rearranging, we obtain

$$\begin{cases} y_i^* = x_i^* \left(\beta_{ivd} - \beta_{ivl}\right) + x_i^* \beta_{ivl} + \mu_i^* \\ \overline{y}_i = \overline{x} \widehat{\beta}_{ivl} + \overline{\mu}_i. \end{cases}$$

Call

$$Y_i^+ = \begin{pmatrix} y_i^* \\ \overline{y}_i \end{pmatrix}, \quad W_i^+ = \begin{pmatrix} x_i^* & x_i^* \\ 0 & \overline{x}_i \end{pmatrix},$$

$$\beta^+ = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_{ivd} - \beta_{ivl} \\ \beta_{ivl} \end{pmatrix}, \quad \mu_i^+ = \begin{pmatrix} \mu_i^* \\ \overline{\mu}_i \end{pmatrix}.$$

The augmented auxiliary model would be

$$Y_i^+ = W_i^+ \beta^+ + \mu_i^+, \qquad i = 1, ..., N. \tag{2}$$

Estimating the model by IV, we obtain directly the variance of the difference of the two estimators in the upper left part of the covariance matrix of β^+. If we now estimate this variance using White's formulae (White, 1984), we get consistent estimators robust to heteroscedasticity and/or dynamic effects within groups. However, in standard econometric packages White's consistent estimators for IV estimators may not be implemented for panel data. In this case, a practical possible solution can be to obtain the IV estimators as OLS estimators on a further transformed model, as it is explained in the next section. This approach is pursued in Appendix 1. The following Lemma is proved.

Lemma 1 Given model (2),

$$\widehat{\beta}_1 = \widehat{\beta}_{ivd} - \widehat{\beta}_{ivl}, \tag{3}$$

$$Var(\widehat{\beta}_1) = Var\left(\widehat{\beta}_{ivd} - \widehat{\beta}_{ivl}\right), \tag{4}$$

An appropriate estimator $\widehat{Var}(\widehat{\beta}_1)$ *consistently estimates* $Var(\widehat{\beta}_1)$. (5)

3 Between groups non spherical disturbances

Let us now turn our attention to cases when the covariance matrix of the auxiliary model is not block diagonal. Consider, for instance, the comparison between a IVD estimator and the WG estimator, that is between an IV estimator and a OLS estimator on the same transformation of the data (deviations from individual time-means).

The formulation of such a test using a standard econometric package is not straightforward. Unlike the cases considered in Section 2, here we do not directly compare *OLS* estimators applied on different orthogonal transformations of the data. In other words, it is not only necessary to manipulate the data according to the different transformations, insert the new variables in an auxiliary regression and then run *OLS* using White (1984) robust standard errors. The procedure also needs to be adjusted. Some preliminaries are needed.

In static models, the most efficient *Generalized Instrumental Variables* estimator is obtained by projecting the variables to be instrumented in the space generated by the instruments. This is a case where the instruments are orthogonal to the initial errors and especially correlated with the initial regressors. It can be shown that, given the properties of the projection matrix, it is equivalent to run *OLS* in a regression where the regressors are the projected variables.[1]

Consider model Then, choose the instrumental matrix, say Z. Project the variables we want to instrument in the space generated by Z

$$\widetilde{X_i^*} = P_Z X_i^*,$$

where

$$P_Z = Z(Z^{'}Z)^{-1}Z^{'}.$$

If we assemble the data in a $NT \times 1$ vector of dependent variables, Y^* and in a $NT \times NK$ matrix of regressors $\widetilde{X^*}$

$$\widehat{\beta}_{ivd} = (\widetilde{X^*}^{'}\widetilde{X^*})^{-1}\widetilde{X^*}Y^*.$$

For the single individual i, construct the system

$$\begin{cases} \widetilde{y_i^*} = \widetilde{x_i^*}\beta + \widetilde{\mu_i^*} \\ y_i^* = x_i^*\beta + \mu_i^*. \end{cases}$$

Estimating by *OLS* the first group of equations, i.e. the ones in levels, we obtain the IVD estimator, i.e. $\widehat{\beta}_{ivd}$. Estimating by *OLS* the second group, i.e. equations in deviations, we obtain the WG estimator, i.e. $\widehat{\beta}_{wg}$. However, the use of an artificial regression, as it is exploited in Sections 2, is not suitable. In Sections 2, because the *between groups* and the *within groups* transformations of the disturbances are orthogonal, the variance covariance matrix in the auxiliary model is block-diagonal. It is then estimated using the White (1984) robust estimators. When other transformations of the data are used, the structure of the variance covariance matrix in the auxiliary regression model can be more complicated. The fact that the equation sets are not orthogonal is not taken into consideration and the White's estimators are not robust to the presence

[1] For further details and an extensive discussion on these issues see Bowden and Turkington (1984).

of inter-groups correlation. The use of a Newey-West robust *OLS* estimator would not help either. The variance covariance matrix exhibits a pattern of cross sectional dependence (i.e. particular form of non stationarity persistent when N goes to infinity) that is not supported by these estimators. Therefore, a consistent estimator for the variance of the difference of the two estimates needs to be constructed step by step. Appendix 2 contains further clarification and the implementations of an appropriate procedure.

4 Conclusions

The paper shows how to estimate the variance of the difference between two panel data estimators that cannot be ranked in terms of efficiency using a standard statistical software. The results are directly applicable in empirical work.

References

[1] Bowden, R.J. and Turkington, D.A. (1984). *Instrumental Variables*, Cambridge University Press.

[2] Davidson, R. and MacKinnon, J.G. (1998). Artificial Regressions, *G.R.E.Q.A.M. Discussion Paper, n. 99a04.*

[3] Hausman, J.A. (1978). Specification Tests in Econometrics, *Econometrica*, 46, 1251-1271.

[4] Lee, M.J. (1996). *Methods of Moments and Semiparametric Econometrics for Limited Dependent Models*, Springer-Verlag Telos.

[5] MacKinnon, J.G. (1992). Model Specification Tests and Artificial Regressions, *Journal of Economic Literature*, 30, 102-146.

[6] O'Brien, R. and Patacchini, E. (2003). Testing the Exogeneity Assumption in Panel Data Models with "Non Classical" Disturbances, *Discussion Paper in Economics and Econometrics, University of Southampton, n. 0302.*

[7] White, H. (1984). *Asymptotic Theory for Econometricians*, Academic Press, New York.

5 Appendix 1

Let us compare an IV estimator on the model transformed according to the *between groups* transformation and an IV estimator on the model transformed according to the *within groups* transformation. The notation follows O'Brien and Patacchini (2003, Appendix 4).

The artificial regression of $Y^* = \begin{bmatrix} HY \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} P_Z HX & P_Z HX \\ 0 & P_Z GX \end{bmatrix}$ gives coefficients $\widehat{\beta}^* = \begin{bmatrix} \widehat{\beta}_{ivd} - \widehat{\beta}_{ivl} \\ \widehat{\beta}_{ivl} \end{bmatrix}$. Results (3) and (4) in Lemma 1 directly follow from the application of Lemma 13 and 14 in O'Brien and Patacchini (2003, Appendix 4). Moreover we use again two orthogonal transformations.

Also in this case

$$Var(Y^*) = \begin{bmatrix} HVar(Y)H' & 0 \\ 0 & GVar(Y)G' \end{bmatrix} = \begin{bmatrix} \frac{\sigma^2}{T}(1+T\theta)I_N & 0 \\ 0 & \sigma^2 I_{N(T-1)} \end{bmatrix}.$$

If now $\widetilde{X} = \begin{bmatrix} P_Z HX & 0 \\ 0 & P_Z GX \end{bmatrix}$,

$$\begin{aligned} Var(\widehat{\beta}^*) &= (X^{*\prime}X^*)^{-1}X^{*\prime}Var(Y^*)X^*(X^{*\prime}X^*)^{-1} \\ &= A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime}. \end{aligned}$$

Next, we calculate this variance by separating the different components.

$$\begin{aligned} \widetilde{X}'Var(Y^*)\widetilde{X} &= \begin{bmatrix} X'H'P_Z' & 0 \\ 0 & X'G'P_Z' \end{bmatrix}\begin{bmatrix} \frac{\sigma^2}{T}(1+T\theta)I_N & 0 \\ 0 & \sigma^2 I_{N(T-1)} \end{bmatrix}\begin{bmatrix} P_Z HX & 0 \\ 0 & P_Z GX \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} X'H'P_Z' & 0 \\ 0 & XG'P_Z' \end{bmatrix}\begin{bmatrix} (\theta+1/T)P_Z HX & 0 \\ 0 & P_Z GX \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} (\theta+1/T)X'H'P_Z'HX & 0 \\ 0 & XG'P_Z'GX \end{bmatrix}. \end{aligned}$$

$$(\widetilde{X}'\widetilde{X})^{-1} = \begin{bmatrix} (X'H'P_Z'HX)^{-1} & 0 \\ 0 & (XG'P_Z'GX)^{-1} \end{bmatrix}.$$

Thus

$$\begin{aligned} &(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1} \\ &= \sigma^2 \begin{bmatrix} (X'H'P_Z'HX)^{-1} & 0 \\ 0 & (XG'P_Z'GX)^{-1} \end{bmatrix} \times \\ &\quad \begin{bmatrix} (\theta+1/T)X'H'P_Z'HX & 0 \\ 0 & XG'P_Z'GX \end{bmatrix}\begin{bmatrix} (X'H'P_Z'HX)^{-1} & 0 \\ 0 & (XG'P_Z'GX)^{-1} \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} (\theta+1/T)(X'H'P_Z'HX)^{-1} & 0 \\ 0 & (XG'P_Z'GX)^{-1} \end{bmatrix} \end{aligned}$$

and

$$
\begin{aligned}
& A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime} \\
= \; & \sigma^2 \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} (\theta+1/T)(X'H'P_Z'HX)^{-1} & 0 \\ 0 & (XG'P_Z'GX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
= \; & \sigma^2 \begin{bmatrix} (\theta+1/T)(X'H'P_Z'HX)^{-1} & -(XG'P_Z'GX)^{-1} \\ 0 & (XG'P_Z'GX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
= \; & \sigma^2 \begin{bmatrix} (\theta+1/T)(X'H'P_Z'HX)^{-1} + (XG'P_Z'GX)^{-1} & -(XG'P_Z'GX)^{-1} \\ -(XG'P_Z'GX)^{-1} & (XG'P_Z'GX)^{-1} \end{bmatrix}. \qquad (6)
\end{aligned}
$$

We now need to find the variance-covariance matrix the artificial regression will assume. This will be proportional to

$$
\begin{aligned}
(X^{*\prime}X^*)^{-1} &= (A'\widetilde{X}'\widetilde{X}A)^{-1} = A^{-1}(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime} \\
&= \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} (X'H'P_Z'HX)^{-1} & 0 \\ 0 & (XG'P_Z'GX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \begin{bmatrix} (X'H'P_Z'HX)^{-1} & -(X'G'P_Z'GX)^{-1} \\ 0 & (X'G'P_Z'GX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \begin{bmatrix} (X'H'P_Z'HX)^{-1} + (X'G'P_Z'GX)^{-1} & -(X'G'P_Z'GX)^{-1} \\ -(X'G'P_Z'GX)^{-1} & (X'G'P_Z'GX)^{-1} \end{bmatrix}. \qquad (7)
\end{aligned}
$$

By comparing (6) with (7) it appears that an artificial regression is a valuable device to estimate a suitable variance-covariance matrix.

We also need to consider the (White) robust *OLS* estimator which uses a consistent estimator of $X^{*\prime}Var(Y^*)X^*$ under the assumption that $Var(Y^*)$ is diagonal.

$$
\begin{aligned}
\widetilde{X}'Var(Y^*)\widetilde{X} &= \begin{bmatrix} X'H'P_Z' & 0 \\ 0 & X'G'P_Z' \end{bmatrix} \begin{bmatrix} \sigma^2\Omega & 0 \\ 0 & \sigma^2\Omega \end{bmatrix} \begin{bmatrix} P_Z HX & 0 \\ 0 & P_Z GX \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} X'P_Z' & 0 \\ 0 & X'G'P_Z' \end{bmatrix} \begin{bmatrix} \Omega & 0 \\ 0 & \Omega \end{bmatrix} \begin{bmatrix} P_Z HX & 0 \\ 0 & P_Z GX \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} X'H'P_Z'\Omega & 0 \\ 0 & X'G'P_Z'\Omega \end{bmatrix} \begin{bmatrix} P_Z HX & 0 \\ 0 & P_Z GX \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} X'H'P_Z'\Omega P_Z HX & 0 \\ 0 & X'G'P_Z'\Omega P_Z GX \end{bmatrix}.
\end{aligned}
$$

Denote for simplicity $\Gamma = X'H'P_Z'\Omega P_Z HX$, $\Pi = X'G'P_Z'\Omega P_Z GX$. Thus

$$
\begin{aligned}
& (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1} \\
= \; & \sigma^2 \begin{bmatrix} (X'H'P_Z'HX)^{-1} & 0 \\ 0 & (XG'P_Z'GX)^{-1} \end{bmatrix} \times \\
& \begin{bmatrix} \Gamma & 0 \\ 0 & \Pi \end{bmatrix} \begin{bmatrix} (X'H'P_Z'HX)^{-1} & 0 \\ 0 & (XG'P_Z'GX)^{-1} \end{bmatrix} \\
= \; & \sigma^2 \begin{bmatrix} (X'H'P_Z'HX)^{-1}\Gamma & 0 \\ 0 & (X'G'P_Z'GX)^{-1}\Pi \end{bmatrix} \begin{bmatrix} (X'H'P_Z'HX)^{-1} & 0 \\ 0 & (X'G'P_Z'GX)^{-1} \end{bmatrix} \\
= \; & \sigma^2 \begin{bmatrix} (X'H'P_Z'HX)^{-1}\Gamma(X'H'P_Z'HX)^{-1} & 0 \\ 0 & (X'G'P_Z'GX)^{-1}\Pi(X'G'P_Z'GX)^{-1} \end{bmatrix}
\end{aligned}
$$

Denote for simplicity $U = (X'H'P_Z'HX)^{-1}\Gamma(X'H'P_Z'HX)^{-1}$,
$V = (X'G'P_Z'GX)^{-1}\Pi(X'G'P_Z'GX)^{-1}$.

$$
\begin{aligned}
& A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime} \\
= \; & \sigma^2 \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
= \; & \sigma^2 \begin{bmatrix} U & -V \\ 0 & V \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} = \sigma^2 \begin{bmatrix} U+V & -V \\ -V & V \end{bmatrix}.
\end{aligned}
$$

The residuals from this regression of $Y^* = \begin{bmatrix} Y \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} P_Z HX & P_Z HX \\ 0 & P_Z GX \end{bmatrix}$ to give coefficients $\widehat{\beta}^* = \begin{bmatrix} \widehat{\beta}_{ivd} - \widehat{\beta}_{ivl} \\ \widehat{\beta}_{ivd} \end{bmatrix}$ can be obtained by stacking those from Y on $P_Z HX$ above those from GY on $P_Z GX$. Similarly to the first artificial regression the first set of equations needs to be scaled by

$$k = \sqrt{T/(1+\theta T)}$$

as otherwise there is no multiple of the residual sum of squares of the artificial regression with expectation σ^2. However, because in this case we are performing an IV estimation by running OLS on a transformed model, the OLS residuals do not provide a consistent estimator of the variance of the initial disturbances. Both in the estimation of θ and in the test statistic, the sum of squares of the residuals has to be calculated using the IV estimate of β and the untransformed right hand side variables. An Hausman test can be calculated by carrying out the artificial regression of $Y^* = \begin{bmatrix} \widehat{k}HY \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} \widehat{k}P_Z HX & \widehat{k}P_Z HX \\ 0 & P_Z GX \end{bmatrix}$ and constructing a Wald test, W, on the first K coefficients, using the following correction:

$$W_{iv} = W_{ols} \frac{\frac{(RSS_A + RSS_B)_{iv}}{[NT-2K]}}{\frac{(RSS_A + RSS_B)_{ols}}{[NT-2K]}} = W_{ols} \frac{(RSS_A + RSS_B)_{iv}}{(RSS_A + RSS_B)_{ols}},$$

where quantities with subscript iv are referred to the initial model and the ones with subscript ols are referred to the transformed model.

6 Appendix 2

Let us compare an IV estimator and an OLS estimator on the model in deviations. In this context, an artificial regression of the type used in Appendix 1 does not help in constructing a test robust for the presence of non spherical errors. In what follows, we explain why it is the case and we indicate an alternative procedure.

Consider the artificial regression of $Y^* = \begin{bmatrix} GY \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} P_Z GX & P_Z GX \\ 0 & GX \end{bmatrix}$. By applying Lemma 13 and 14 in O'Brien and Patacchini (2003, Appendix 4), we get that $\widehat{\beta}^* = \begin{bmatrix} \widehat{\beta}_{ivd} - \widehat{\beta}_{wg} \\ \widehat{\beta}_{wg} \end{bmatrix}$. The disturbances $\varepsilon^* = \begin{bmatrix} P_Z Gu \\ Gu \end{bmatrix}$ have a covariance matrix $E(\varepsilon^* \varepsilon^{*\prime}) = \sigma^2 \begin{bmatrix} P_Z & P_Z \\ P_Z & I_{N(T-1)} \end{bmatrix}$, as $GG' = I_{N(T-1)}$.

In this case, the transformations of the data used in the two sets of equations are not orthogonal and $Var(Y^*)$ is not diagonal. We have

$$Var(Y^*) = \begin{bmatrix} GVar(Y)G' & GVar(Y)G' \\ GVar(Y)G' & GVar(Y)G' \end{bmatrix} = \begin{bmatrix} \sigma^2 I_{N(T-1)} & \sigma^2 I_{N(T-1)} \\ \sigma^2 I_{N(T-1)} & \sigma^2 I_{N(T-1)} \end{bmatrix}.$$

If now $\widetilde{X} = \begin{bmatrix} P_Z GX & 0 \\ 0 & GX \end{bmatrix}$,

$$\begin{aligned} Var(\widehat{\beta}^*) &= (X^{*\prime}X^*)^{-1}X^{*\prime}Var(Y^*)X^*(X^{*\prime}X^*)^{-1} \\ &= A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime}. \end{aligned}$$

Next, we calculate this variance by separating the different components.

$$\widetilde{X}'Var(Y^*)\widetilde{X} = \begin{bmatrix} X'G'P_Z' & 0 \\ 0 & X'G' \end{bmatrix} \begin{bmatrix} \sigma^2 I_{N(T-1)} & \sigma^2 I_{N(T-1)} \\ \sigma^2 I_{N(T-1)} & \sigma^2 I_{N(T-1)} \end{bmatrix} \begin{bmatrix} P_Z GX & 0 \\ 0 & GX \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} X'G'P_Z' & X'G'P_Z' \\ X'G' & X'G' \end{bmatrix} \begin{bmatrix} P_Z GX & 0 \\ 0 & GX \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} X'G'P_Z'GX & X'G'P_Z'GX \\ X'G'P_Z'GX & X'QX \end{bmatrix}.$$

$$(\widetilde{X}'\widetilde{X})^{-1} = \begin{bmatrix} (X'G'P_Z'GX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix}.$$

Thus

$$(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1}$$

$$= \sigma^2 \begin{bmatrix} (X'G'P_Z'GX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \times$$

$$\begin{bmatrix} X'G'P_Z'GX & X'G'P_Z'GX \\ X'G'P_Z'GX & X'QX \end{bmatrix} \begin{bmatrix} (X'G'P_Z'GX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} I & I \\ (X'QX)^{-1}[X'G'P_Z'GX] & I \end{bmatrix} \begin{bmatrix} (X'G'P_Z'GX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} (X'G'P_Z'GX)^{-1} & (X'QX)^{-1} \\ (X'QX)^{-1} & (X'QX)^{-1} \end{bmatrix}$$

and

$$\begin{aligned}
& A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime} \\
= \; & \sigma^2 \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} (X'G'P_Z'GX)^{-1} & (X'QX)^{-1} \\ (X'QX)^{-1} & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \qquad (8) \\
= \; & \sigma^2 \begin{bmatrix} (X'G'P_Z'GX)^{-1} - (X'QX)^{-1} & 0 \\ (X'QX)^{-1} & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
= \; & \sigma^2 \begin{bmatrix} (X'G'P_Z'GX)^{-1} - (X'QX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix}. \qquad (9)
\end{aligned}$$

If we run the artificial regression, the postulated variance-covariance matrix is different. It will be proportional to

$$\begin{aligned}
(X^{*\prime}X^*)^{-1} &= (A'\widetilde{X}'\widetilde{X}A)^{-1} = A^{-1}(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime} \\
&= \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} (X'G'P_Z'GX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \begin{bmatrix} (X'G'P_Z'GX)^{-1} & -(X'QX)^{-1} \\ 0 & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \begin{bmatrix} (X'G'P_Z'GX)^{-1} + (X'QX)^{-1} & -(X'QX)^{-1} \\ -(X'QX)^{-1} & (X'QX)^{-1} \end{bmatrix}.
\end{aligned}$$

The fact that the equation sets in the auxiliary regression constructed are not orthogonal is not taken into consideration. A wrong answer will also come from the White's estimators. They are not robust to the presence of inter-groups correlation. The use of a Newey-West robust *OLS* estimator would not help either. The variance covariance matrix exhibits a pattern of cross sectional dependence (i.e. particular form of non stationarity persistent when N goes to infinity) that is not supported by these estimators. Therefore, a consistent estimator for the variance of the difference of the two estimates (upper left part of matrix (9)) needs to be constructed step by step. We need to recover the matrices involved and a consistent estimate of σ. Recall that for the first set we are performing an *IV* estimation by running *OLS* on a transformed model. Therefore it is known that the *OLS* sum of squares are not a consistent estimate of the variance of the initial disturbances because the transformed model produces a non spherical variance-covariance matrix. The sum of squares of the residuals coming from the initial model with the *IV* estimator should be used instead.

However, notice that

$$\widehat{\varepsilon}_{iv} = y - X\widehat{\beta}_{iv}$$

can be written as

$$\begin{aligned}\widehat{\varepsilon}_{iv} &= y - X\widehat{\beta}_{ols} + X\widehat{\beta}_{ols} - X\widehat{\beta}_{iv} \\ &= \widehat{\varepsilon}_{ols} + X\left(\widehat{\beta}_{ols} - \widehat{\beta}_{iv}\right)\end{aligned}$$

and therefore

$$\widehat{\varepsilon}'_{iv}\widehat{\varepsilon}_{iv} = \widehat{\varepsilon}'_{ols}\widehat{\varepsilon}_{ols} + \left(\widehat{\beta}_{ols} - \widehat{\beta}_{iv}\right)' X'X\left(\widehat{\beta}_{ols} - \widehat{\beta}_{iv}\right).$$

The sum of squares of the residuals coming from the initial model with the IV estimator is equal to the OLS sum of squares plus a function of the contrast between the two estimators, which is what we want to test eventually. This contaminates the variance estimate. Therefore, in order to get a consistent estimator of the variance we can rely only on the second set.

We run OLS on the first set of equations and use White robust standard errors. They produce a consistent estimator of $X'Var(Y)X$ under the assumption that $Var(Y) = \sigma^2\Omega$, a block diagonal matrix.

We get

$$\begin{aligned}X'Var(GY)X &= X'G'P'_Z\left[\sigma^2\Omega\right]P_Z GX \\ &= \sigma^2\left(X'G'P'_Z\Omega P_Z GX\right).\end{aligned}$$

So

$$\begin{aligned}&(X'X)^{-1}X'Var(Y)X'(X'X)^{-1} \\ =\ &\sigma_1^2\left(X'G'P'_Z GX\right)^{-1}\left(X'G'P'_Z\Omega P_Z GX\right)\left(X'G'P'_Z GX\right)^{-1}.\end{aligned}$$

In order to get the matrix of interest, we will divide the estimate of this variance by the obtained $\widehat{\sigma}_1^2$.

Denote $\Psi = \left(X'G'P'_Z GX\right)^{-1}\left(X'G'P'_Z\Omega P_Z GX\right)\left(X'G'P'_Z GX\right)^{-1}$.

Similarly, run OLS on the first set of equations and use White robust standard errors.

We get

$$\begin{aligned}X'Var(GY)X &= X'G'\left[\sigma^2\Omega\right]GX \\ &= \sigma^2\left(X'G'\Omega GX\right).\end{aligned}$$

So

$$\begin{aligned}&(X'X)^{-1}X'Var(Y)X'(X'X)^{-1} \\ =\ &\sigma_2^2\left(X'QX\right)^{-1}\left(X'G'\Omega GX\right)\left(X'QX\right)^{-1}.\end{aligned}$$

Denote $\Theta = \left(X'QX\right)^{-1}\left(X'G'\Omega GX\right)\left(X'QX\right)^{-1}$.

A robust and consistent estimator of the precision matrix in the Wald test is[2]

$$\left[\widehat{Var}\left(\widehat{\beta}_{ivd} - \widehat{\beta}_{wg}\right)\right]^{-1} = \widehat{\sigma}_2^2\left(\Psi - \Theta\right).$$

[2]Note that the precision matrix may not always be positive definite in finite samples.