The Estimation of Simultaneous Equation Models under Conditional Heteroscedasticity

EMMA M. IGLESIAS* and GARRY D.A. PHILLIPS
University of Alicante Cardiff University

This version: April 2004. Preliminary and to be completed

Abstract

In this paper we extend the setting analysed in Hahn and Hausman (2002a) by allowing for conditionally heteroscedastic disturbances. We start by considering the type of conditional variance-covariance matrices proposed by Engle and Kroner (1995) and we show that, when we impose a GARCH specification in the structural model, some conditions are needed to have a GARCH process of the same order in the reduced form equations. Later, we propose a modified-2SLS and a modified-3SLS procedures where the conditional heteroscedasticity is taken into account, that are more asymptotically efficient than the traditional 2SLS and 3SLS estimators. We recommend to use these modified-2SLS and 3SLS procedures in practice instead of alternative estimators like LIML/FIML, where the non-existence of moments leads to extreme values (in case we are interested in the structural form). We show theoretically and with simulation that in some occasions 2SLS, 3SLS and our proposed 2SLS and 3SLS procedures can have very severe biases, and we present the bias correction mechanisms to apply in practice.

1 Introduction

Following the seminal work of Engle (1982), a large number of papers have dealt with conditionally heteroscedastic disturbances in many different settings. Most of the theory has been developed in a univariate framework, although more recently multivariate models have been explored. In relation to simultaneous equations, Baba, Engle, Kraft and Kroner (1991), Harmon (1988), and Engle and Kroner (1995) have introduced the theoretical framework of simultaneous equation models with conditional

*Corresponding author. Dpt. Fundamentos del Análisis Económico. University of Alicante. Campus de San Vicente, 03080, Alicante, Spain. Fax: +34 96 5903898. E-mail address: emiglesias@merlin.fae.ua.es.

heteroscedastic disturbances, although the theoretical approach is still not well developed. In this paper we provide a theoretical and simulation study of the behaviour of 2SLS, LIML/FIML and 3SLS estimators in the context of simultaneous equations with ARCH disturbances in the framework of Hausman (1983), Phillips (2003) and Hahn and Hausman (2002a, 2002b, 2003). We will compare the behaviour of 2SLS and 3SLS with alternative 2SLS and 3SLS estimators that take account of the ARCH structure and which have better asymptotic and finite sample properties. We show as well that LIML can have problems because of the non-existence of moments, whereby the modified-2SLS and 3SLS estimation procedures proposed in this paper are preferred for practical application. As stated in Hahn and Hausman (2002a) in relation to LIML, "these results should be a caution about using LIML estimates...without further investigation or specification tests in a given empirical problem". This is a problem that specially has been reported in the presence of weak instruments in the literature (see Hahn, Hausman and Kuersteiner (2002)). The same type of conclusion is found in this paper in the context of conditional heteroscedastic disturbances, although we find the same problem even already without weak instruments.

The structure of the paper is as follows. Section 2 examines how, in a very simple framework, it is possible to allow for conditional heteroscedasticity within the context of a 2SLS estimation procedure following which we develop a modified procedure which is asymptotically more efficient. The improved efficiency of the modified procedure is then confirmed in a set of Monte Carlo simulations. The simulations show that the small sample biases in both 2SLS and in the modified-2SLS estimator that we propose can be very severe in some circumnstances, and we consider a bias-correction mechanism for practical application. In Section 3 we present the results of LIML estimation. We find simulation evidence of the problem of non-existence of moments in LIML, and recommend in practice the use of our modified-2SLS procedure. Section 4 examines a more general simultaneous system with conditional heteroscedastic disturbances, where we extend our approach to 3SLS and again find that a modified version is more asymptotically efficient. Section 5 explores the context of weak instruments in this setting, and finally, Section 6 concludes.

2 Efficient 2SLS estimation of a simultaneous equation system with the presence of conditional heteroscedastic disturbances

Engle and Kroner (1995) noted that a simultaneous equation system can be consistently estimated with 2SLS or 3SLS while ignoring the conditionally heteroscedastic structure, although they do not analyse the theoretical properties of the estimators. We proceed now to analyse 2SLS in a simple framework. We will consider first 2SLS where we do not take into account the ARCH structure and then a modified 2SLS ($2SLS_M$) which makes use of the conditional heteroscedastic characteristics of the disturbances to estimate the system more efficiently. We shall, initially, follow the set up employed by Hausman (1983), Hahn and Hausman (2002a, 2002b, 2003) and Phillips (2003) which analysed a very simple model

$$\begin{aligned} y_{1t} &= \beta_1 y_{2t} + \varepsilon_{1t} \\ y_{2t} &= \beta_2 y_{1t} + x_t' \gamma_1 + \varepsilon_{2t} \\ y_{2t} &= x_t' \pi_2 + v_{2t} \end{aligned} \tag{1}$$

where the third equation corresponds to the reduced form, π_2 is K$\times$1 with K$\geq$ 2 and T is the sample size. In this case, only the first of the equations is identified; in fact it is overidentified of order at least 2. The variables in x_t are assumed to be strongly exogenous and bounded. The main novelty of this paper is that we are going to allow for conditional heteroscedasticity in the structural disturbances according to the following

$$h_{11t} = E(\varepsilon_{1t}^2 | I_{t-1}), h_{22t} = E(\varepsilon_{2t}^2 | I_{t-1}), h_{12t} = E(\varepsilon_{1t}\varepsilon_{2t} | I_{t-1}) \tag{2}$$

Before proceeding with our analysis we examined the characteristics of the reduced form disturbances when the structural disturbances are conditionally heteroscedastic. For this simple case we find in the next Proposition that when the structural distrubances follow a multivariate-ARCH process, the reduced form disturbances may also be a multivariate-ARCH process of the same order but only under quite strict conditions. In particular, the variance parameters in the ARCH processes must be the same. This appears to run counter to Proposition 3.1 in Engle and Kroner (1995) which asserts that the result holds generally.

Proposition 2.1. If $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ is a multivariate conditional heteroscedastic process in (1), and $v_t = B^{-1}\varepsilon_t$ is the reduced form disturbance vector where $B = \begin{pmatrix} 1 & -\beta_1 \\ -\beta_2 & 1 \end{pmatrix}$ then, under appropriate conditions, v_t may follow a multivariate conditional heteroscedastic process of the same order as ε_t, but the result is not true generally.

Proof. Given in Appendix 1. ■

We return now to the analysis of the system given in (1). While we generally give proofs of our theoretical results in the appendices, it will be appropriate here to motivate our approach by considering how efficiency can be gained by taking account of the conditional heteroscedasticity in the context of $2SLS$ estimation.

Consider the first equation of (1), and to take account of the conditional heteroscedasticity in this equation we transform it to

$$\frac{y_{1t}}{\sqrt{h_{11t}}} = \beta_1 \frac{y_{2t}}{\sqrt{h_{11t}}} + \frac{\varepsilon_{1t}}{\sqrt{h_{11t}}} \tag{3}$$

where now the disturbances have mean zero and variance unity and they continue to be serially uncorrelated. In the usual $2SLS$ procedure the endogenous regressor is replaced by its predicted value obtained from regressing the endogenous regressor on all the predetermined variables. In this case the endogenous regressor has been standardised by $\sqrt{h_{11t}}$ and so the corresponding predicted value comes from the regression

$$\frac{y_{2t}}{\sqrt{h_{11t}}} = \frac{x_t'}{\sqrt{h_{11t}}}\pi_2 + \frac{v_{2t}}{\sqrt{h_{11t}}} \tag{4}$$

Writing the predicted value as $(\frac{\widehat{y}_{2t}}{\sqrt{h_{11t}}}) = \frac{x_t'}{\sqrt{h_{11t}}}\widehat{\pi}_2$, where $\frac{y_{2t}}{\sqrt{h_{11t}}} = \frac{x_t'}{\sqrt{h_{11t}}}\widehat{\pi}_2 + \frac{\widehat{v}_{2t}}{\sqrt{h_{11t}}}$ and the residuals are orthogonal to the predicted values, we find that a modified-$2SLS$ estimator is given by

$$\begin{aligned} \beta_{1,2SLS_M} &= \sum\left(\frac{\widehat{y}_{2t}y_{1t}}{h_{11t}}\right)/\sum(\frac{\widehat{y}_{2t}}{\sqrt{h_{11t}}})^2 \\ &= \beta_1 + \sum\left(\frac{\widehat{y}_{2t}\varepsilon_{1t}}{h_{11t}}\right)/\sum(\frac{\tilde{y}_{2t}}{\sqrt{h_{11t}}})^2 \\ &= \beta_1 + \sum\frac{x_t'\widehat{\pi}_2\varepsilon_{1t}}{h_{11t}}/\sum(\frac{x_t'\widehat{\pi}_2}{\sqrt{h_{11t}}})^2 \end{aligned} \tag{5}$$

where it is easy to show that this estimator is consistent and its asymptotic variance is

$$\begin{aligned} avar\sqrt{T}(\beta_{1,2SLS_M} - \beta_1) &= \frac{1}{E(\frac{1}{\sqrt{h_{11t}}})}\lim(\sum T^{-1}(\pi_2'x_tx_t'\pi_2))^{-1} \\ &= \frac{1}{E(\frac{1}{\sqrt{h_{11t}}})}\,(\pi_2'\sum_{xx}\pi_2))^{-1} \end{aligned} \tag{6}$$

where we have used $lim\sum T^{-1}x_tx_t' = \sum_{xx}$, a finite positive definite matrix.

The usual $2SLS$ which we write as β_1^* has asymptotic variance given by

$$avar\sqrt{T}(\beta_1^* - \beta_1) = \sigma_{11}(\pi_2'\sum_{xx}\pi_2))^{-1} \tag{7}$$

so that the asymptotic relative efficiency of the modified $2SLS$ estimator is given by the ratio of the asymptotic variances $\frac{1}{E(\frac{1}{\sqrt{h_{11t}}})}/\sigma_{11}$. Noting that $E(\frac{1}{\sqrt{h_{11t}}}) > \frac{1}{E(\sqrt{h_{11t}})} > \frac{1}{\sqrt{\sigma_{11}}}$, it follows that $\frac{1}{E(\frac{1}{\sqrt{h_{11t}}})}/\sigma_{11} < 1$, thus demonstrating the advantage, in terms of asymptotic efficiency, of accounting for the conditional heteroscedasticity. In practice the modified estimator discussed here is infeasible since the conditional variances are unknown and must be estimated. However with appropriate assumptions the above asymptotics will still hold. In the context of the model in (1) and the multivariate

$ARCH$ process in (2), the operational $2SLS_M$ estimator is then given by the following procedure

STEP 1: Obtain the residuals by running a first round of the traditional 2SLS without taking into account the ARCH effects.

STEP 2: Regress these residuals in a multivariate ARCH system to get the estimates of $\widehat{h}_{11t}$.

STEP 3: Regress $\frac{y_{2t}}{\sqrt{\widehat{h}_{11t}}}$ on $\frac{x_t}{\sqrt{\widehat{h}_{11t}}}$ to find $\frac{\widehat{y}_{2t}}{\sqrt{\widehat{h}_{11t}}} = \frac{x_t\widehat{\pi}_2}{\sqrt{\widehat{h}_{11t}}}$ which is orthogonal to $\frac{\widehat{v}_{2t}}{\sqrt{\widehat{h}_{11t}}}$.

STEP 4: Put $\frac{y_{1t}}{\sqrt{\widehat{h}_{11t}}} = \beta_1\frac{\widehat{y}_{2t}}{\sqrt{\widehat{h}_{11t}}} + \frac{\varepsilon_{1t}}{\sqrt{\widehat{h}_{11t}}} + \frac{\widehat{v}_{2t}}{\sqrt{\widehat{h}_{11t}}}$, where $\sum_{t=1}^{T}\frac{\widehat{y}_{2t}}{\sqrt{\widehat{h}_{11t}}}\frac{\widehat{v}_{2t}}{\sqrt{\widehat{h}_{11t}}} = 0$, and regress $\frac{y_{1t}}{\sqrt{\widehat{h}_{11t}}}$ on $\frac{\widehat{y}_{2t}}{\sqrt{\widehat{h}_{11t}}}$ to obtain

$$\beta_{1,2SLS_M} = \left[\sum_{t=1}^{T}\left(\frac{\widehat{y}_{2t}}{\sqrt{\widehat{h}_{11t}}}\right)^2\right]^{-1}\sum_{t=1}^{T}\frac{\widehat{y}_{2t}}{\sqrt{\widehat{h}_{11t}}}\frac{y_{1t}}{\sqrt{\widehat{h}_{11t}}} =$$

$$\beta_1 + \sum_{t=1}^{T}\left[\frac{\widehat{y}_{2t}\varepsilon_{1t}}{\widehat{h}_{11t}} / \frac{\widehat{y}_{2t}^2}{\widehat{h}_{11t}}\right]$$

Being straightforward to show that it is consistent: $p\lim\beta_{1,2SLS_M} = \beta_1 + \frac{p\lim\frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{y}_{2t}\varepsilon_{1t}}{\widehat{h}_{11t}}}{p\lim\frac{1}{T}\sum_{t=1}^{T}\frac{\widehat{y}_{2t}^2}{\widehat{h}_{11t}}}$.

While the asymptotic relative efficiency of the 2SLS$_M$ procedure has been demonstrated in the simplest case, the result extends directly to cases where the equation has more endogenous and exogenous variables which is discussed again in Section 4.

2.1 Small sample properties of 2SLS and modified-2SLS: evidence from simulations in a simple model

In relation to the finite sample biases, the modified-2SLS procedure proposed in this paper and the standard 2SLS procedure, are both biased. We give below simulation evidence of how both procedures, can yield estimators with very severe biases in some circumstances, and bias-correction is often necessary. It is already well known in the literature that the 2SLS is biased. In relation to the traditional 2SLS, the Nagar (1959) bias approximation for 2SLS in the simple model where only the first equation is identified, and where the disturbances are normally, independently and identically distributed, specialises to

$$E(\beta_1^* - \beta_1) = tr\left[(P_X - P_W - 1)\right] \frac{1}{\pi_2' X' X \pi_2} \frac{\sigma_{11}\beta_2 + \sigma_{12}}{1 - \beta_1\beta_2} \quad + o(T^{-1}). \tag{8}$$

where β_1 and β_2 are given in (1), and P_X is the projection matrix based on the matrix X and P_W is the projection matrix based on W, the non-stochastic part of y_2. Its trace is equal to the number of variables in X, i.e. the number of exogenous variables including the constant. In this case what is called W is just a vector so $trace(P_W)$ is just equal to one and $tr\left[(P_X - P_W - 1)\right]$ is just equal to the order of overidentification minus one.

The above bias approximation results as a corollary of the analysis given in Phillips (2003), where it was shown that it is sufficient for the structural disturbances to have Gauss Markov properties unconditionally for the bias approximation to be valid. Since the ARCH disturbances are unconditionally Gauss Markov, we can assert that the above bias approximation is valid for the model here also.

While it is straightforward to bias correct the usual $2SLS$ estimator based upon an estimate of the Nagar bias, we do not have a bias approximation for the modified estimator, $2SLS_{\hat{M}}$,and so an alternative approach is necessary. Bias correction by the bootstrap is possible and here we set out how the method might be used. A later version of this paper will explore this this further and, in particular, present some Monte Carlo evidence of the value of the approach.

Returning to the simple model (1) of section 2, the $2SLS_M$ was given as

$$\beta_{1,2SLS_{\hat{M}}} = \left[\sum_{t=1}^{T}\left(\frac{\widehat{y}_{2t}}{\sqrt{\widehat{h}_{11t}}}\right)^2\right]^{-1} \sum_{t=1}^{T} \frac{\widehat{y}_{2t}}{\sqrt{\widehat{h}_{11t}}} \frac{y_{1t}}{\sqrt{\widehat{h}_{11t}}}. \quad \text{Substituting}$$

$$\frac{y_{1t}}{\sqrt{\widehat{h}_{11t}}} = \beta_1 \frac{\widehat{y}_{2t}}{\sqrt{\widehat{h}_{11t}}} + \frac{\varepsilon_{1t}}{\sqrt{\widehat{h}_{11t}}} + \frac{\widehat{v}_{2t}}{\sqrt{\widehat{h}_{11t}}}$$

and noting that the vector of components containing $\frac{\widehat{y}_{2t}}{\sqrt{\widehat{h}_{11t}}}$ is orthogonal to the vector containing the components of $\frac{\widehat{v}_{2t}}{\sqrt{\widehat{h}_{11t}}}$,we find that

$$\beta_{1,2SLS_{\hat{M}}} - \beta_1 = \left[\sum_{t=1}^{T} \frac{\widehat{y}_{2t}^2}{\widehat{h}_{11t}}\right]^{-1} \sum_{t=1}^{T} \frac{\widehat{y}_{2t}}{\sqrt{\widehat{h}_{11t}}} \frac{\varepsilon_{1t}}{\sqrt{\widehat{h}_{11t}}}$$

so that the bias of $\beta_{1,2SLS_{\hat{M}}}$ is given by the expected value of this expression.

To apply the bootstrap, we first estimate the structural equation by the usual $2SLS$ method and retain the reduced form and structural equation residuals. Resampling with replacement from these residuals and constructing the pseudo data making use of the original parameter estimates will provide a bootstrapped bias correction for $2SLS$. To find a bootstrap bias correction for the modified estimator proceeds along exactly similar lines except that in each sample of pseudo data the $\beta_{1,2SLS_{\hat{M}}}$ estimate is obtained and the average of these subtracted from the original estimate provides the bias correction. This approach will be explored in more detail in the next version of the paper.

We proceed now to present some simulation results which confirm that the modified-2SLS procedure is more efficient than the traditional 2SLS procedure, but which also show that both methods can be severely biased in some circumnstances and that bias-correction in both cases may be necessary. Table 1 provides simulations for a sample of size 100 based on 5000 replications, and the structure we consider is of the form

$$YB + X\Gamma + \varepsilon = 0$$

where $B = \begin{pmatrix} -1 & 0.267 \\ 0.222 & -1 \end{pmatrix}$ and $\Gamma = \begin{pmatrix} 0 & 0 & 0 \\ 4.40 & 0.74 & 0.13 \end{pmatrix}'$.

The matrix X contains a first column of ones, while the other two exogenous variables correspond to normal random variables that have been generated with a mean of zero and variance 10. The model has been estimated by 2SLS and 2SLS$_M$.

To represent the behaviour of the disturbances in the structural system we have selected, for reasons of operational simplicity, the model of Wong and Li (1997) that follows the structure

$$E\left(\varepsilon_{1t}^2/I_{t-1}\right) = \alpha_0 + \alpha_1\varepsilon_{1t-1}^2 + \alpha_2\varepsilon_{2t-1}^2$$

$$E\left(\varepsilon_{2t}^2/I_{t-1}\right) = \gamma_0 + \gamma_1\varepsilon_{1t-1}^2 + \gamma_2\varepsilon_{2t-1}^2$$

In our simulations we also provide the bias approximation results of the formula given in Phillips for the traditional 2SLS procedure. In the Wong and Li model, in which the disturbances are contemporaneously and serially uncorrelated and homoscedastic, the bias approximation will imply $\sigma_{12} = 0$ in the formula given in (8). Thus the bias will then depend directly on β_2 and σ_{11}. Results are given in Table 1 below.

Table 1: Simulation results for 2SLS and 2SLS$_M$

	2SLS ignoring ARCH		2SLS without ignoring ARCH	
	bias($\widehat{\beta}$)	s.e.($\widehat{\beta}$)	bias($\widehat{\beta}$)	s.e.($\widehat{\beta}$)
$\alpha_0 = 0.81, \alpha_1 = 0.25, \alpha_2 = 0.49$	0.002	0.038	0.000	0.029
$\gamma_0 = 0.64, \gamma_1 = \gamma_2 = 0.49$	(0.001)			
$\alpha_0 = 9, \alpha_1 = 0.25, \alpha_2 = 0.49$	0.014	0.101	-0.012	0.095
$\gamma_0 = 0.64, \gamma_1 = \gamma_2 = 0.49$	(0.004)			
$\alpha_0 = 0.81, \alpha_1 = 0.49, \alpha_2 = 0.49$	0.005	0.053	0.002	0.048
$\gamma_0 = 0.64, \gamma_1 = \gamma_2 = 0.49$	(0.004)			
$\alpha_0 = 144, \alpha_1 = 0.25, \alpha_2 = 0.49$	0.130	0.309	-0.099	0.292
$\gamma_0 = 0.64, \gamma_1 = \gamma_2 = 0.49$	(0.058)			

In brackets we provide the numerical value of the Nagar approximation for the bias. The first interesting result to note, is that indeed the modified-2SLS procedure is more efficient than the traditional 2SLS procedure. In addition, the $2SLS_M$ estimator has a smaller absolute bias while both procedures can have very severe biases especially when the unconditional variance of the disturbance of the first equation is large. Then, bias correction will be necessary. If the researcher uses 2SLS without taking account the ARCH system, then the Nagar bias approximation should be helpful (although it does not account for more than about half of the bias in some scenarios). In case the researcher follows our suggested procedure, Table 1 shows that, although it has less bias than the traditional 2SLS, bias correction is still necessary and we recommend to apply it through the boostrap.

3 LIML estimation of a simultaneous equation system with conditional heteroscedasticity

In the setting that we have been analysing so far, where only the first of the equations is identified, 2SLS and 3SLS provide the same result. Engle and Kroner (1995) propose to estimate the system more efficiently as well through full information maximum likelihood or an instrumental variable estimator. In this case, because in our context the second of the equations is not identified, FIML will be equal to LIML. In this section we proceed now to consider this estimation method.

Table 2 provides results based on 5000 replications and a sample size of 100, for LIML for the case where we do not take account of the ARCH effects

Table 2: Simulation results for LIML

	LIML ignoring ARCH	
	bias($\widehat{\beta}$)	s.e.($\widehat{\beta}$)
$\alpha_0 = 0.81, \alpha_1 = 0.25, \alpha_2 = 0.49$ $\gamma_0 = 0.64, \gamma_1 = \gamma_2 = 0.49$	-0.001	0.035
$\alpha_0 = 9, \alpha_1 = 0.25, \alpha_2 = 0.49$ $\gamma_0 = 0.64, \gamma_1 = \gamma_2 = 0.49$	-0.010	0.091
$\alpha_0 = 0.81, \alpha_1 = 0.49, \alpha_2 = 0.49$ $\gamma_0 = 0.64, \gamma_1 = \gamma_2 = 0.49$	-0.005	0.053
$\alpha_0 = 144, \alpha_1 = 0.25, \alpha_2 = 0.49$ $\gamma_0 = 0.64, \gamma_1 = \gamma_2 = 0.49$	-2.812	0.355

Care is needed in interpreting these results since it is unclear that estimator moments exist. It is well known that in the classical simultaneous model with normal disturbances, finite sample LIML estimators do not have moments of any order and a similar non-existence of moments problem may exist here. Indeed extreme values were present in the simulations especially for the fourth structure examined. If we were to consider LIML estimation taking account of the presence of ARCH effects explicitly in the LIML procedure, this seems likely to produce a Quasi-LIML estimator where the moments would not exist either (considering ARCH effects imply even fatter tails for the disturbances than under normality). That is why in this paper, when conditional heteroscedasticity is present and we are interested in the structural parameters, we recommend to use in practice of a 2SLS procedure that takes into account the ARCH effects rather than LIML. Because this type of 2SLS uses the disturbance standardised, it has moments even when the disturbance presents ARCH effects. It would exist as well the possibility of estimating the reduced form coefficients using a Full Information Maximum Likelihood (FIML) approach after standardising the reduced form equations individually. However, in this paper we are interested in the structural-form coefficients and, as it is seen in Table 2, sometimes the estimates that can be obtained through FIML can be heavily affected by the non-existence of moments mainly when the variance of the first disturbance is quite large (a problem which is also documented in Hahn and Hausman (2002a) for the case of unconditional correlation when they do not allow for conditional heteroscedasticity).

4 Modified 2SLS and 3SLS estimation of a general simultaneous equation system

So far, we have carried out the analysis in the context of (1) to make easier the interpretation of the analysis. In this section we develop the theoretical approach in a more general setting such as

$$\begin{aligned} y_{1t} &= \beta_1 y_{2t} + x_{1t}'\gamma_1 + \varepsilon_{1t} \\ y_{2t} &= \beta_2 y_{1t} + x_{2t}'\gamma_2 + \varepsilon_{2t} \end{aligned} \tag{9}$$

In this context, the structural form can be alternatively written

$$\begin{aligned} y_{1t} &= z_{1t}'\alpha_1 + \varepsilon_{1t} \\ y_{2t} &= z_{2t}'\alpha_2 + \varepsilon_{2t} \end{aligned} \tag{10}$$

where
$z_{1t}' = (y_{2t} : x_1') \, , z_{2t}' = (y_{1t} : x_2') \, .$

We shall assume that each equation omits at least two exogenous variables and so is overidentified at least of order 2. As before, we assume that

$$h_{11t} = E(\varepsilon_{1t}^2 | I_{t-1}), h_{22t} = E(\varepsilon_{2t}^2 | I_{t-1}), h_{12t} = E(\varepsilon_{1t}\varepsilon_{2t} | I_{t-1}) \tag{11}$$

and that the structural disturbances are unconditionally Gauss Markov. Although this is again a simple two-equation model it proves to be completely appropriate for our purposes since all our results can be extended directly to a general simultaneous equation model containing G equations.

Before we proceed to examine the modified -2SLS estimator, we first consider the estimation of the reduced form parameters. The reduced form equations will be

$$\begin{aligned} y_{1t} &= x_t'\pi_1 + v_{1t} \\ y_{2t} &= x_t'\pi_2 + v_{2t} \end{aligned} \tag{12}$$

In obtaining the 2SLS estimator for the parameters of the first structural equation we require to estimate the reduced form equation for y_2.To find a modified estimator of the vector of reduced form pararmeters π_2,we rewrite the equation as

$$\frac{y_{2t}}{\sqrt{h_{2t}^v}} = \frac{x'\pi_2}{\sqrt{h_{2t}^v}} + \frac{v_{2t}}{\sqrt{h_{2t}^v}}, \quad t = 1, 2, ..., T. \tag{13}$$

where the variables have been standardised by the conditional standard deviation of the disturbance v_{2t} and not by $\sqrt{h_{11t}}$ as is required in the modified $2SLS$ procedure. The modified-OLS estimator here is more asymptotically efficient than OLS which is summed up in the following:

Theorem 4.1. The modified-OLS estimator of the reduced form parameter vector π_2 in (13) , is asymptotically more efficient than the OLS estimator which ignores the presence of conditional heteroscedasticity.
Proof. Given in Appendix 2. ■

We now consider the modified-2SLS procedure in the context of the model (9) and (11) for which the first stage regression is conducted in (12). This estimator is referred to as 2SLS_M.The fact that estimation is improved by taking the conditional heteroscedasticity into account is summed up in the following

Theorem 4.2. Under the simultaneous equation system defined in (9) and (11), 2SLS_M is more asymptotically efficient than 2SLS.
Proof. Given in Appendix 3. ■

Note that when using the modified-$2SLS$ estimator the first round regression is not based on equation (13) but on

$$\frac{y_{2t}}{\sqrt{h_{11t}}} = \frac{x'\pi_2}{\sqrt{h_{11t}}} + \frac{v_{2t}}{\sqrt{h_{11t}}}, \quad t = 1, 2, ..., T$$

where, in order to have orthogonality between the residuals and the predicted value of $\frac{y_{2t}}{\sqrt{h_{11t}}}$ which enters the second stage regression, the variables are standardised by the conditional standard deviation of the structural disturbance and not the reduced form disturbance. Although the resulting estimator of π_2 is not explicitly used, it is of interest to compare its asymptotic efficiency with that which results in Theorem 4.2. We do this in the following

Theorem 4.3. If the alternative modified-OLS estimator of π_2 which results from the regression in (13) is used to construct an alternative modified $2SLS$ estimator, the resulting estimator may be more or less asymptotically efficient than the estimator in Theorem 4.2.
Proof. Given in Appendix 4. ■

We know from the standard literature that 3SLS is always more efficient than 2SLS when the equations are overidentified and the disturbances are contemporaneously correlated. Thus, in the model of this section, 3SLS is more asymptotically efficient than 2SLS and so might be preferred. However, we shall see that it too is less asymptotically efficient than a modified-3SLS ($3SLS_M$) procedure. This $3SLS_M$ procedure will imply in practical applications to follow a similar procedure than the traditional 3SLS, but where again we standardise by the conditional variances of the structural disturbances.

First consider again the structural equations

$$\begin{aligned} y_{1t} &= \beta_1 y_{2t} + x'_{1t}\gamma_1 + \varepsilon_{1t} \\ y_{2t} &= \beta_2 y_{1t} + x'_{2t}\gamma_2 + \varepsilon_{2t}, \ t = 1, 2, ..., T. \end{aligned} \tag{14}$$

We shall write the system as

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} Z_1 & 0 \\ 0 & Z_2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

where $Z_i = (y_i : X_i)$, $\alpha_i = (\ \beta_i \quad \gamma'_i\ )'$, $i = 1.2$.

Premultiplying by X',the matrix which contains all the exogenius variables, yields the system

$$\begin{pmatrix} X'y_1 \\ X'y_2 \end{pmatrix} = \begin{pmatrix} X'Z_1 & 0 \\ 0 & X'Z_2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} X'\varepsilon_1 \\ X'\varepsilon_2 \end{pmatrix}$$

where the covariance matrix of the transformed disturbances is

$$E \begin{pmatrix} X'\varepsilon_1 \\ X'\varepsilon_2 \end{pmatrix} \begin{pmatrix} X'\varepsilon_1 \\ X'\varepsilon_2 \end{pmatrix}' = \begin{pmatrix} \sigma_{11}X'X & \sigma_{12}X'X \\ \sigma_{21}X'X & \sigma_{22}X'X \end{pmatrix}$$

$= \Sigma \otimes (X'X)$

where $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$.

Applying GLS to this system yields the 3SLS-GLS estimator

$$\begin{pmatrix} \sigma^{11} Z_1' P_X Z_1 & \sigma^{12} Z_1' P_X Z_2 \\ \sigma^{21} Z_2' P_X Z_1 & \sigma^{22} Z_2' P_X Z_2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma^{11} Z_1' P_X y_1 + \sigma^{12} Z_1' P_X y_2 \\ \sigma^{22} Z_2' P_X y_2 + \sigma^{21} Z_2' P_X y_1 \end{pmatrix} \tag{15}$$

To obtain the modified 3SLS estimator we first define two diagonal matrices given by

$$\Lambda_1 = \begin{pmatrix} \frac{1}{\sqrt{h_{111}}} & 0 & ... & 0 \\ 0 & \frac{1}{\sqrt{h_{112}}} & ... & 0 \\ ... & ... & ... & 0 \\ 0 & 0 & ... & \frac{1}{\sqrt{h_{11T}}} \end{pmatrix}, \Lambda_2 = \begin{pmatrix} \frac{1}{\sqrt{h_{221}}} & 0 & ... & 0 \\ 0 & \frac{1}{\sqrt{h_{222}}} & ... & 0 \\ ... & ... & ... & 0 \\ 0 & 0 & ... & \frac{1}{\sqrt{h_{22T}}} \end{pmatrix}$$

These matrices are used to standardise the variables in the system so that the system becomes

$$\begin{pmatrix} \Lambda_1 y_1 \\ \Lambda_2 y_2 \end{pmatrix} = \begin{pmatrix} \Lambda_1 Z_1 & 0 \\ 0 & \Lambda_2 Z_2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \Lambda_1 \varepsilon_1 \\ \Lambda_2 \varepsilon_2 \end{pmatrix}$$

Premultiplying by X'yields

$$\begin{pmatrix} X'\Lambda_1 y_1 \\ X'\Lambda_2 y_2 \end{pmatrix} = \begin{pmatrix} X'\Lambda_1 Z_1 & 0 \\ 0 & X'\Lambda_2 Z_2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} X'\Lambda_1 \varepsilon_1 \\ X'\Lambda_2 \varepsilon_2 \end{pmatrix}$$

where the covariance matrix of the transformed disturbances is

$$E \begin{pmatrix} X'\Lambda_1 \varepsilon_1 \\ X'\Lambda_2 \varepsilon_2 \end{pmatrix} \begin{pmatrix} X'\Lambda_1 \varepsilon_1 \\ X'\Lambda_2 \varepsilon_2 \end{pmatrix}' = \begin{pmatrix} X'X & \rho_{12} X'X \\ \rho_{21} X'X & X'X \end{pmatrix}$$

$$= \Sigma_M \otimes (X'X)$$

where

$$\Sigma_M = \begin{pmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{pmatrix} \quad \text{with } \rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}. \tag{16}$$

Applying GLS to this transformed system will yield the modified-3SLS-GLS estimator

$$\begin{pmatrix} \tilde{\alpha}_1 \\ \tilde{\alpha}_2 \end{pmatrix} = \begin{pmatrix} Z_1'\Lambda_1 P_x \Lambda_1 Z_1 & -\rho_{12} Z_1'\Lambda_1 P_x \Lambda_2 Z_2 \\ -\rho_{21} Z_2'\Lambda_2 P_x \Lambda_1 Z_1 & Z_2'\Lambda_2 P_x \Lambda_2 Z_2 \end{pmatrix}^{-1} \begin{pmatrix} Z_1'\Lambda_1 P_x \Lambda_1 y_1 - \rho_{12} Z_1'\Lambda_1 P_x \Lambda_2 y_2 \\ Z_2'\Lambda_2 P_x \Lambda_2 y_2 - \rho_{12} Z_2'\Lambda_2 P_x \Lambda_1 y_1 \end{pmatrix}$$

We may now state the following:

Theorem 4.4. Under the simultaneous equation system defined in (9) and (11), $3SLS_M$ is more asymptotically efficient than $3SLS$.
Proof. Given in Appendix 5. ■

The results in Theorems 4.2 and 4.4 are given in the context of the estimators $2SLS_M$ and $3SLS_M$ both of which are non-operational since the standardising conditional standard deviations are unknown. However, the conditional standard deviations can be consistently estimated from the residuals obtained following first round estimation so that operational versions are readily found. These operational estimators will have the same asymptotic distribution as the $2SLS_M$ and $3SLS_M$ counterparts. This matter will be considered further in the next version of the paper.

5 Simultaneous equations and weak instruments under conditional heteroscedasticity

To be completed in a later version of the paper.

6 Conclusions

In this paper we have studied simultaneous equation systems and how 2SLS and 3SLS behave in this framework. First we have shown, that linear combinations of ARCH processes do not produce always ARCH processes of the same order, unless some conditions are hold. We have also proposed modified 2SLS and 3SLS procedures that are more asymptotically efficient than the traditional procedures. When the researcher is interested in estimating the structural parameters, we recommend to use our modified procedures instead of LIML (or FIML) where the existence of extreme values can produce misleading results in practice (due to the non-existence of moments, and even more under the context of conditional heteroscedasticy where the tails are fatter than in the regular case). We have also showed through Monte Carlo simulations how all the procedures can produce important biases, mainly when the disturbances are very volatile, and we provide bias mechanisms to apply in practice.

7 Appendices

Appendix 1
Proof. of Proposition 2.1.

In this appendix we show that if the structural disturbances follow a multivariate-ARCH(1) process, the reduced form disturbances may also follow a multivariate-ARCH(1) process but only under strict conditions. To show this we suppose that the disturbances in (1) follow, for example, a diagonal representation (for simplicity, but without loss of generality)

$$h_{11t} = E(\varepsilon_{1t}^2|I_{t-1}) = \alpha_0 + \alpha_1\varepsilon_{1t-1}^2, h_{22t} = E(\varepsilon_{2t}^2|I_{t-1}) = \theta_0 + \theta_1\varepsilon_{2t-1}^2,$$

$$h_{12t} = E(\varepsilon_{1t}\varepsilon_{2t}|I_{t-1}) = \lambda_1 + \lambda_2\varepsilon_{1,t-1}\varepsilon_{2,t-1}$$

Note that v_{2t}, the reduced form disturbance in the second equation, is defined by

$$v_{2t} = \frac{(\beta_2\varepsilon_{1t} + \varepsilon_{2t})}{1 - \beta_1\beta_2}, \beta_1\beta_2 \neq 1.$$

Also $E(v_{2t}^2) = \frac{\beta_2^2\sigma_{11} + 2\beta_2\sigma_{12} + \sigma_{22}}{(1-\beta_1\beta_2)^2}$, while the conditional variance is given by

$$\begin{aligned} E(v_{2t}^2|I_{t-1}) &= \frac{\beta_2^2}{(1-\beta_1\beta_2)^2}(\alpha_0 + \alpha_1\varepsilon_{1,t-1}^2) + \frac{2\beta_2}{(1-\beta_1\beta_2)^2}(\lambda_1 + \lambda_2\varepsilon_{1,t-1}\varepsilon_{2,t-1}) \\ &\quad + \frac{1}{(1-\beta_1\beta_2)^2}(\theta_0 + \theta_1\varepsilon_{2,t-1}^2) \\ &= \frac{\beta_2^2}{(1-\beta_1\beta_2)^2}h_{1t} + \frac{2\beta_2}{(1-\beta_1\beta_2)^2}h_{12t} + \frac{1}{(1-\beta_1\beta_2)^2}h_{2t} \end{aligned}$$

Next we have $v_{2t-1}^2 = \frac{\beta_2^2\varepsilon_{1t-1}^2 + 2\beta_2\varepsilon_{1,t-1}\varepsilon_{2,t-1} + \varepsilon_{2,t-1}^2}{(1-\beta_1\beta_2)^2}$, from which it is apparent that it is not possible to write $E(v_{2t}^2|I_{t-1}) = \phi_1 + \phi_2 v_{2t-1}^2$ for some ϕ_1, ϕ_2 unless restrictions are placed on the original $ARCH(1)$ processes. In particular, for v_{2t} to follow an $ARCH(1)$ process of the usual kind the component $ARCH$ processes will have to have the same variance parameter. Clearly this is a severe restriction to impose, and this proves the proposition.

Similarly we may show that the 2×1 vector $v = \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}$ has a conditional covariance matrix given by

$E(vv'|I_{t-1}) = \begin{bmatrix} h_{1t}^v & h_{12t}^v \\ h_{21t}^v & h_{2t}^v \end{bmatrix}$, where

$$h_{1t}^v = \frac{\beta_1^2}{(1-\beta_1\beta_2)^2} h_{2t} + \frac{2\beta_1}{(1-\beta_1\beta_2)^2} h_{12t} + \frac{1}{(1-\beta_1\beta_2)^2} h_{1t}$$

$$h_{12t}^v = h_{21t}^v = \beta_2 h_{1t} + (1+\beta_1\beta_2) h_{12t} + \beta_1 h_{2t}$$

$$h_{2t}^v = \frac{\beta_2^2}{(1-\beta_1\beta_2)^2} h_{1t} + \frac{2\beta_2}{(1-\beta_1\beta_2)^2} h_{12t} + \frac{1}{(1-\beta_1\beta_2)^2} h_{2t}$$

■

Appendix 2

Proof. of Theorem 4.1.

We shall write rewrite the equation in (12) by putting $\frac{y_{2t}}{\sqrt{h_{2t}^v}} = y_{2t}^*$, $\frac{x'\pi_2}{\sqrt{h_{2t}^v}} = x_t^*$ and $\frac{v_{2t}}{\sqrt{h_{2t}^v}} = v_{2t}^*$.With T observations we may write the regression as

$y_2^* = X^*\pi_2 + v_2^*$.

Then the GLS (because we have standardised) estimator for $\pi_2$is given by

$\hat{\pi}_2 = ((X^*)'X^*)^{-1}(X^*)'y_1^*$

$= \pi_2 + ((X^*)'X^*)^{-1}(X^*)'v_1^*$

from which $\sqrt{T}(\hat{\pi}_2 - \pi_2)$ has an asymptotic covariance matrix given by

$$\lim_{T\to\infty} (T^{-1}(X^*)'X^*)^{-1} = \frac{1}{E(\frac{1}{h_{2t}^v})} \lim_{T\to\infty} (T^{-1}X'X)^{-1} \tag{17}$$

The asymptotic covariance matrix for the OLS estimator is :

$\lim_{T\to\infty} \sigma_{22}(T^{-1}X'X)^{-1}$(being σ_{22} the unconditional variance). Hence the relative efficiency is $\frac{\sigma_{22}}{E(\frac{1}{h_{2t}^v})}$ (or $\frac{E(\frac{1}{h_{1t}^v})}{\sigma_{11}}$). We know from Jensen's inequality that $E(\frac{1}{h_{2t}^v}) > \frac{1}{E(h_{2t}^v)} = \frac{1}{\sigma_{11}}$ so that $\frac{\sigma_{11}}{E(\frac{1}{h_{2t}^v})} > 1$, and so the result is proved. A similar result will hold for $\hat{\pi}_1$.

If in addition, the disturbances are jointly symmetric, it is possible to prove straightforwardly that the modfiied-OLS reduced form parameter estimator is unbiased. ■

Appendix 3

Proof. of Theorem 4.2.

In the structural system defined by (9) and (11), let's define $\alpha^* = (\alpha_1^*, \alpha_2^*)'$ to be the 2SLS estimator. Then

$\alpha_1^* = \left(Z_1'X\left(X'X\right)^{-1}X'Z_1\right)^{-1} Z_1'X\left(X'X\right)^{-1}X'y_1$

$\alpha_2^* = \left(Z_2'X\left(X'X\right)^{-1}X'Z_2\right)^{-1} Z_2'X\left(X'X\right)^{-1}X'y_2$

Analysing the distrubution of $\sqrt{T}\left(\alpha_1^* - \alpha_1\right)$, the asymptotic covariance matrices are given by (where σ_{11} and σ_{22} are the two unconditional variances of the structural disturbances)

$avar\left(\sqrt{T}(\alpha_1^* - \alpha_1)\right) = \sigma_{11} p\lim T\left(Z_1'X\left(X'X\right)^{-1}X'Z_1\right)^{-1}$

$avar\left(\sqrt{T}(\alpha_2^* - \alpha_2)\right) = \sigma_{22} p\lim T\left(Z_2'X\left(X'X\right)^{-1}X'Z_2\right)^{-1}$

In the case of our modified-2SLS procedure, let's define $\widetilde{\alpha} = (\widetilde{\alpha}_1, \widetilde{\alpha}_2)$ to be the modified-2SLS estimator. Then, put

$$\Lambda_1 = \begin{pmatrix} \frac{1}{\sqrt{h_{111}}} & 0 & ... & 0 \\ 0 & \frac{1}{\sqrt{h_{112}}} & ... & 0 \\ ... & ... & ... & 0 \\ 0 & 0 & ... & \frac{1}{\sqrt{h_{11T}}} \end{pmatrix}, \Lambda_2 = \begin{pmatrix} \frac{1}{\sqrt{h_{221}}} & 0 & ... & 0 \\ 0 & \frac{1}{\sqrt{h_{222}}} & ... & 0 \\ ... & ... & ... & 0 \\ 0 & 0 & ... & \frac{1}{\sqrt{h_{22T}}} \end{pmatrix}$$

We may show that

$\widetilde{\alpha}_1 = \left(Z_1'\Lambda_1 P_X \Lambda_1 Z_1\right)^{-1} Z_1'\Lambda_1 P_X \Lambda_1 y_1$

$\widetilde{\alpha}_2 = \left(Z_2'\Lambda_2 P_X \Lambda_2 Z_2\right)^{-1} Z_2'\Lambda_2 P_X \Lambda_2 y_2$

where we have written $P_X = X(X'X)^{-1}X'$. The asymptotic covariance matrix of $\tilde{\alpha}_1$ is $avar\left(\sqrt{T}(\tilde{\alpha}_1 - \alpha_1)\right) = p\lim T\left(Z_1'\Lambda_1 P_X \Lambda_1 Z_1\right)^{-1} = \frac{1}{\left(E\left(\frac{1}{\sqrt{h_{11t}}}\right)\right)^2} p\lim\left(\frac{1}{T} Z_1'P_x Z_1\right)^{-1}$

Using Jensen's inequality

$\left(E\left(\frac{1}{\sqrt{h_{11t}}}\right)\right)^2 \geq \frac{1}{\left(E\left(\sqrt{h_{11t}}\right)\right)^2}$

$E\left(\left(\sqrt{h_{11t}}\right)^2\right) = \sigma_{11} \leq \left(E\left(\sqrt{h_{11t}}\right)\right)^2 \Rightarrow \frac{1}{\left(E\left(\sqrt{h_{11t}}\right)\right)^2} \leq \sigma_{11}.$

Thus we have proved that this non-operational $2SLS_M$ is more asymptotically efficient than 2SLS. The same would hold for $\widetilde{\alpha}_2$. ■

Appendix 4

Proof. of Theorem 4.3.

Suppose we now use the modified OLS estimator of $\hat{\pi}_2$ to construct the modified $2SLS$ estimator. We now have the equation:

$y_1^{**} = \beta_1 \hat{y}_2^{**} + \varepsilon_1^{**} + \beta_1 \hat{v}_2^{**},$

The usual situation does not apply here: $\hat{y}_2^{**} = X^{**}\hat{\pi}_2$ is not orthogonal to the second component of the error term $\beta_1 \hat{v}_2$. However, the implied 2SLS estimator will still be

$\widetilde{\beta}_1 = [(\hat{\pi}_2)'(X^{**})'X^{**}\hat{\pi}_2\}^{-1}(\hat{\pi}_2)'(X^{**})'y_1^{**}$

$= \beta_1 + [(\hat{\pi}_2)'(X^{**})'X^{**}\hat{\pi}_2\}^{-1}(\hat{\pi}_2)'(X^{**})'(\varepsilon_1^{**} + \beta_1 \hat{v}_2^{**}) =$

$\beta_1 + [(\hat{\pi}_2)'(X^{**})'X^{**}\hat{\pi}_2\}^{-1}(\hat{\pi}_2)'(X^{**})'\varepsilon_1^{**} + [(\hat{\pi}_2)'(X^{**})'X^{**}\hat{\pi}_2\}^{-1}(\hat{\pi}_2)'(X^{**})'\beta_1\hat{v}_2^{**})$
Then as $T \to \infty$
$\sqrt{T}(\ \tilde{\beta}_1 -\ \beta_1\ ) \sim\ \ [(\pi_2)'\frac{1}{T}(X^{**})'X^{**}\pi_2]^{-1}(\pi_2)'T^{-1/2}(X^{**})'(\varepsilon_1^{**} + \beta_1 v_2^{**}),$
which has asymptotic covariance matrix given by

$$AsyVar(\sqrt{T}(\ \tilde{\beta}_1 -\ \beta_1\ )) =\ \ var(\varepsilon_{1t}^{**} + \beta_1 v_{2t}^{**})\frac{1}{(E(\frac{1}{h_{11t}^{w}}))}\ \lim_{T\to\infty}[(\pi_2)'\frac{1}{T}(X'X)\pi_2]^{-1}$$

which may be more or less asymptotically efficient than the usual $2SLS_M$ depending on whether or not $var(\varepsilon_{1t}^{**} + \beta_1 v_{2t}^{**}) \leqslant 1$. ■

Appendix 5

Proof. of Theorem 4.4.

In the structural system defined by (9) and (11), let $\alpha^{**} = (\alpha_1^{**}, \alpha_2^{**})'$ to be the 3SLS estimator. Then, the asymptotic covariance matrix will be given by

$$avar\left(\sqrt{T}\begin{pmatrix}\alpha_1^{**}-\alpha_1\\ \alpha_2^{**}-\alpha_2\end{pmatrix}\right) = p\lim T\begin{pmatrix}\sigma^{11}Z_1'P_XZ_1 & \sigma^{12}Z_1'P_XZ_2\\ \sigma^{21}Z_2'P_XZ_1 & \sigma^{22}Z_2'P_XZ_2\end{pmatrix}^{-1}.$$

$$= p\lim T\begin{pmatrix}\frac{1}{1-\rho_{12}^2}\frac{1}{\sigma_{11}}Z_1'P_xZ_1 & -\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}\frac{1}{1-\rho_{12}^2}Z_1'P_xZ_2\\ -\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}\frac{1}{1-\rho_{12}^2}Z_2'P_xZ_1 & \frac{1}{1-\rho_{12}^2}\frac{1}{\sigma_{22}}Z_2'P_xZ_2\end{pmatrix}^{-1}$$

$$= (1-\rho_{12}^2)\quad p\lim T\begin{pmatrix}\frac{1}{\sigma_{11}}Z_1'P_xZ_1 & -\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}Z_1'P_xZ_2\\ -\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}Z_2'P_xZ_1 & \frac{1}{\sigma_{22}}Z_2'P_xZ_2\end{pmatrix}^{-1}$$

on which makes use of the result that the unconditional variance/covariance matrix can be written as

$$\Sigma^{-1} = \begin{pmatrix}\sigma^{11} & \sigma^{12}\\ \sigma^{21} & \sigma^{22}\end{pmatrix} = \frac{1}{1-\rho_{12}^2}\begin{pmatrix}\frac{1}{\sigma_{11}} & -\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}\\ -\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}} & \frac{1}{\sigma_{22}}\end{pmatrix}$$

Here we have defined the unconditional correlation coefficient as ρ_{12}. On the other hand, if we define the modified-3SLS estimator by $\tilde{\tilde{\alpha}} = \left(\tilde{\tilde{\alpha}}_1, \tilde{\tilde{\alpha}}_2\right)$, then, the asymptotic covariance matrix is given by

$$avar\left(\sqrt{T}\begin{pmatrix}\tilde{\tilde{\alpha}}_1-\alpha_1\\ \tilde{\tilde{\alpha}}_2-\alpha_2\end{pmatrix}\right) = p\lim T\begin{pmatrix}\frac{1}{1-\rho_{12}^2}Z_1'\Lambda_1P_X\Lambda_1Z_1 & \frac{-\rho_{12}}{1-\rho_{12}^2}Z_1'\Lambda_1P_X\Lambda_2Z_2\\ \frac{-\rho_{21}}{1-\rho_{12}^2}Z_2'\Lambda_2P_X\Lambda_1Z_1 & \frac{1}{1-\rho_{12}^2}Z_2'\Lambda_2P_X\Lambda_2Z_2\end{pmatrix}^{-1} =$$

$$(1-\rho_{12}^2)p\lim T\begin{pmatrix}\left(E\left(\frac{1}{\sqrt{h_{11t}}}\right)\right)^2 Z_1'P_XZ_1 & -\rho_{12}E\left(\frac{1}{\sqrt{h_{11t}}}\right)E\left(\frac{1}{\sqrt{h_{22t}}}\right)Z_1'P_XZ_2\\ -\rho_{12}E\left(\frac{1}{\sqrt{h_{11t}}}\right)E\left(\frac{1}{\sqrt{h_{22t}}}\right)Z_2'P_XZ_1 & \left(E\left(\frac{1}{\sqrt{h_{22t}}}\right)\right)^2 Z_2'P_XZ_2\end{pmatrix}^{-1}$$

By repeated application of Jensen's inequality, we may show that $avar\left(\sqrt{T}\left(\alpha^{**}-\alpha\right)\right) - avar\left(\sqrt{T}\left(\tilde{\tilde{\alpha}}-\alpha\right)\right)$ is positive semi-definite, thus proving that $3SLS_M$ is asymptotically more efficient than $3SLS$. ■

References

[1] Baba, Y., R. F. Engle, D. F. Kraft and K. F. Kroner (1991), Multivariate Simultaneous Generalised ARCH, University of California, San Diego: Department of Economics, Discussion Paper No. 89-57.

[2] Engle, R. F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50, 987-1007.

[3] Engle, R. F. and K. F. Kroner (1995), Multivariate Simultaneous Generalised ARCH, *Econometric Theory* 11, 122-150.

[4] J. Hahn and J. A. Hausman (2002a), "A New Specification Test for the Validity of Instrumental Variables", *Econometrica* 70, 163-189.

[5] J. Hahn and J. A. Hausman (2002b), "Notes on Bias in Estimators for Simultaneous Equation Models", *Economics Letters* 75, 2, 237-41.

[6] J. Hahn and J. A. Hausman (2003), "Weak Instruments: Diagnosis and Cures in Empirical Econometrics", *American Economic Review,* 93, 2, 118-125.

[7] J. Hahn, J. A. Hausman and G. Kuersteiner (2002), "Estimation with Weak Instruments: Accuracy of Higher Order Bias and MSE Approximations", MIT Working Paper.

[8] Harmon, R. (1988), The simultaneous Equations Model with Generalised Autoregressive Conditional Heteroscedasticity: the SEM-GARCH Model, Washington D. C.: Board of Governors of the Federal Reserve System, International Finance Discussion Papers, No. 322.

[9] Hausman, J. A. (1983), "Specification and Estimation of Simultaneous Equation Models" in Griliches, Zvi and Intriligator, Michael, eds., *Handbook of Econometrics*, Volume 1, Amsterdam : North Holland.

[10] Nagar, A. L. (1959), The Bias and Moment Matrix of the General k-class Estimators of the Parameters in Simultaneous Equations. *Econometrica* 27, 575-95.

[11] Phillips, G. D. A. (2003), "Nagar-type moment approximations in simultaneous equation models: some further results. Paper presented at Contributions to Econometric Theory: Conference in Memory of Michael Magdalinos, Athens November 2003.

[12] Wong, H. and W. K. Li (1997), On a Multivariate Conditional Heteroscedastic Model, *Biometrika* 84, 1, 111-123.