ROBUST WALD TESTS IN SUR SYSTEMS WITH ADDING UP RESTRICTIONS: AN ALGEBRAIC APPROACH TO PROOFS OF INVARIANCE

SURAJIT RAY, B. RAVIKUMAR, AND N. EUGENE SAVIN

Abstract. In this paper, we examine the robust Wald test statistic for SUR systems with adding up restrictions where the same explanatory variables are present in all equations and where heteroskedasticity and/or autocorrelation of unknown forms may be present. For this case, the coefficients are usually estimated by least squares, equation by equation. For testing the typical hypotheses of interest, we show that the robust Wald statistic, i.e., the statistic based on the heteroskedasticity and autocorrelation consistent covariance matrix estimator, is invariant to the equation deleted. Our proof of invariance is algebraic and does not rely on parametric assumptions or on the knowledge of the covariance matrix of disturbances. Furthermore, the adding-up restrictions we consider are of a general form: the weighted sum of the dependent variables adds up to one of the explanatory variables, not necessarily a constant. We illustrate our results using the Capital Asset Pricing Model.

1. Introduction

For SUR systems with adding-up restrictions, it is well known that the covariance matrix of disturbances is singular. The usual approach to hypothesis testing in such cases is to construct the relevant test statistics after deleting an equation. A common application of this approach is in the context of complete demand systems where the sum of expenditure shares must equal one. Barten (1969) considered the maximum likelihood estimation of such a system of equations with independent and identical normal disturbance vectors. He proved that the value of the likelihood function and, hence, the maximum likelihood estimates of the parameters are invariant to the equation deleted. This, in turn, implies that the value of the likelihood ratio statistic for testing linear restrictions on the coefficients is invariant to the equation deleted. Similarly, McGuire, Farley, Lucas and Ring (1969), Powell (1969) and Theil (1971) considered the Generalized Least Squares (GLS) estimation of a system of demand equations. Under the assumption that the covariance matrix of the stacked disturbance vector

Date: January 1998.

Key words and phrases. SUR System, Adding up, Wald test, Heteroskedasticity, Autocorrelation.

is known, they showed that the GLS estimator and the corresponding quadratic form are invariant to the equation deleted. Estimation and testing have been extended to SUR systems with specific forms of heteroskedasticity and/or autocorrelations; see, for instance, Mandy and Martins-Filho (1993) and Berndt and Savin (1975).

In practice, the likelihood function and/or the covariance matrix of the stacked disturbance vector are usually unknown. Similarly, the functional form of heteroskedasticity and/or autocorrelations is also unknown. In this paper, we consider SUR systems with adding-up restrictions where the same explanatory variables are present in all equations and where heteroskedasticity and/or autocorrelation of unknown forms may be present. For this case, the coefficients are usually estimated by least squares, equation by equation. For testing the typical hypotheses of interest, we show that the robust Wald statistic, i.e., the statistic based on the heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator, is invariant to the equation deleted. Our proof of invariance is algebraic and does not rely on parametric assumptions as in Barten (1969) or on the knowledge of covariance matrix as in Powell (1969). Furthermore, the adding-up restrictions we consider are more general than Barten's. As in Powell, the weighted sum of the dependent variables in this paper adds up to one of the explanatory variables, not necessarily a constant.

We illustrate our results using the Sharpe-Lintner Capital Asset Pricing Model (CAPM). In the CAPM, the return on the market portfolio is a weighted sum of returns on the individual assets. Thus, one can think of the CAPM as a SUR system with an adding-up restriction. In conventional tests of the CAPM, the null hypothesis is that the vector of intercepts is zero; see Campbell, Lo and MacKinlay (1997). The singularity problem (associated with the covariance matrix of disturbances) is usually ignored in testing this hypothesis. In our illustration, we use the HAC Wald statistic after deleting an equation.

The organization of the paper is as follows. In section 2, we develop our results for the Classical SUR system. In the process, we provide an alternative proof of Barten's result that does not rely on normal distributions. In section 3, we consider a SUR system with heteroskedastic disturbances of an unknown form. We construct the Wald test statistic for this case using the White (1980) heteroskedasticity consistent (HC) covariance matrix estimator and show that the HC Wald statistic is invariant to the equation deleted. In section 4, we consider SUR systems with heteroskedastic and autocorrelated distrubances. The CAPM example is in section 5 and the concluding remarks are contained in section 6.

2. Classical SUR System

Consider the following system of equations:

$$
\begin{array}{rcl}
y_{1t} & = & \beta_{11}x_{1t} + \beta_{12}x_{2t} + \cdots + \beta_{1k}x_{kt} + \varepsilon_{1t} \\
y_{2t} & = & \beta_{21}x_{1t} + \beta_{22}x_{2t} + \cdots + \beta_{2k}x_{kt} + \varepsilon_{2t} \\
 & \vdots & \\
y_{nt} & = & \beta_{n1}x_{1t} + \beta_{n2}x_{2t} + \cdots + \beta_{nk}x_{kt} + \varepsilon_{nt}
\end{array}
\tag{2.1}
$$

Define $\varepsilon_{.t} = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{nt})^{'}$, an $(n \times 1)$ vector of disturbances at time period t. Assume that the disturbances vector are i.i.d. and satisfy the following:

$$
\left.\begin{array}{rcl}
E(\varepsilon_{it}) & = & 0 \\
E(\varepsilon_{it}\varepsilon_{jt}^{'}) & = & \sigma_{ij}
\end{array}\right\} \quad \forall\, t.
\tag{2.2}
$$

The assumptions in (2.2) can also be written as

$$
\left.\begin{array}{rcl}
E(\varepsilon_{.t}) & = & 0 \\
E(\varepsilon_{.t}\varepsilon_{.t}^{'}) & = & \Sigma^*
\end{array}\right\} \quad \forall\, t,
\tag{2.3}
$$

where $\Sigma^* = \{\sigma_{ij}\}$, $i, j = 1, 2, \ldots, n$, is a $(n \times n)$ contemporaneous covariance matrix. The classical SUR system where the same explanatory variables appear in all the equations is specified by (2.1) and (2.3). See Greene (section 15.4, (1997)), for details.

The above system of equations can be written in the following compact form

$$
\begin{array}{rcl}
y_1 & = & X\beta_1 + \varepsilon_1 \\
y_2 & = & X\beta_2 + \varepsilon_2 \\
 & \vdots & \\
y_n & = & X\beta_n + \varepsilon_n
\end{array}
\tag{2.4}
$$

where y_i is an $(T \times 1)$ vector, X an $(T \times k)$ matrix of explanatory variables, $\beta_i = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{ik})^{'}$ an $(k \times 1)$ vector of parameters and ε_i an $(T \times 1)$ vector of disturbances.

The system of equations in (2.4) can be stacked as

$$
y = Z\gamma + \varepsilon
\tag{2.5}
$$

where $y = (y_1, y_2, \ldots, y_n)^{'}$ is an $(nT \times 1)$ column vector of the dependent variables, γ is an $(nk \times 1)$ vector of parameters, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^{'}$ is a $(nT \times 1)$ column vector of disturbances and Z is an $(kT \times kT)$ block diagonal matrix,

$$
\begin{array}{rcl}
Z & = & \begin{bmatrix} X & 0 & \ldots & 0 \\ 0 & X & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & X \end{bmatrix} \\
 & = & I_n \otimes X
\end{array}
$$

where I_n is an identity matrix of order n. The covariance matrix, Σ, for the above SUR model is

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1n} \\ \Sigma_{21} & \Sigma_{22} & \dots & \Sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{n1} & \Sigma_{n2} & \dots & \Sigma_{nn} \end{bmatrix}$$

where the ijth block, Σ_{ij}, is

$$\Sigma_{ij} = \begin{bmatrix} \sigma_{ij} & 0 & \dots & 0 \\ 0 & \sigma_{ij} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{ij} \end{bmatrix}. \tag{2.6}$$

Hence, $\Sigma_{ij} = \sigma_{ij} I_T$, where I_T is a identity matrix of order T. From equations (2.3) and (2.6) the following holds for the above SUR model:

$$\begin{aligned} E(\varepsilon) &= 0, \\ E(\varepsilon\varepsilon') &= \Sigma, \\ &= \Sigma^* \otimes I_T. \end{aligned} \tag{2.7}$$

Define $\beta_{.j} = (\beta_{1j}, \beta_{2j}, \dots, \beta_{nj})'$, an $(n \times 1)$ column vector of parameters corresponding to the jth explanatory variable. The null hypothesis of interest is

$$H_0 \colon \beta_{.j} = 0, \tag{2.8}$$

or equivalently $H_0 \colon r\gamma = 0$, where $r = I_n \otimes R_j'$ is a selection matrix, and R_j is k dimensional column vector with 1 in the jth position and zeros elsewhere. The alternative hypothesis is $H_1 \colon \beta_{.j} \neq 0$.

The OLS estimate of γ and the variance of this estimate are given by

$$\begin{aligned} \hat{\gamma} &= (Z'Z)^{-1}Z'y, \\ V(\hat{\gamma}) &= (Z'Z)^{-1}Z'\Sigma Z(Z'Z)^{-1}. \end{aligned} \tag{2.9}$$

Substituting $Z = I_n \otimes X$ and $\Sigma = \Sigma^* \otimes I_T$ in the expression for the variance of $\hat{\gamma}$ and simplifying, we obtain

$$V(\hat{\gamma}) = \Sigma^* \otimes (X'X)^{-1}. \tag{2.10}$$

A well known procedure to test the above hypothesis (2.8) is to use a Wald test statistic. Assuming that the covariance matrix Σ is known, the Wald statistic is given by

$$J_1 = \hat{\beta}_{.j}' \left[rV(\hat{\gamma})r' \right]^{-1} \hat{\beta}_{.j} \tag{2.11}$$

where $\hat{\beta}_{.j}$ is the least-squares estimate of $\beta_{.j}$. Substituting $r = I_n \otimes R_j'$ and the expression for the variance from (2.10), the Wald statistic can be written as

$$\begin{aligned} J_1 &= \hat{\beta}'_{.j}\left[(I_n \otimes R'_j)(\Sigma^* \otimes (X'X)^{-1})(I_n \otimes R_j)\right]^{-1}\hat{\beta}_{.j} \\ &= \hat{\beta}'_{.j}\left[\Sigma^* \otimes R'_J(X'X)^{-1}R_j\right]^{-1}\hat{\beta}_{.j} \\ &= \hat{\beta}'_{.j}\left[\Sigma^* \otimes c\right]^{-1}\hat{\beta}_{.j} \end{aligned}$$

where c is the jth diagonal element of $(X'X)^{-1}$. Hence, the Wald statistic can be re-written as

$$J_1 = c^{-1}\hat{\beta}'_{.j}\,(\Sigma^*)^{-1}\,\hat{\beta}_{.j}. \tag{2.12}$$

Under the null hypothesis, J_1 has an asymptotic chi-square distribution with n degrees of freedom. In general Σ^* is unknown and in practice a consistent estimator for Σ^* is used in (2.12) for testing H_0.

Adding up

Now assume that the above system satisfies the adding up condition.

$$\omega' y_{.t} = x_{1t} \quad \forall\, t, \tag{2.13}$$

where $\omega = (\omega_1, \omega_2, \ldots, \omega_n)'$ is a weight vector such that $\sum_{j=1}^{n}\omega_j = 1$. We have assumed without loss of generality that $\omega' y_{.t}$ adds up to the first variable x_{1t}. Equation (2.13) implies the following restrictions on the parameters and disturbances:

$$\begin{aligned} \omega'\beta_{.1} &= 1, \\ \omega'\beta_{.j} &= 0, \quad j = 2, 3, \ldots, k\ , \end{aligned} \tag{2.14}$$

and,

$$\omega'\varepsilon_{.t} = 0 \quad \forall\, t. \tag{2.15}$$

Equation (2.15) implies that the ε_{it}'s, $i = 1, 2, \ldots, n$, are linearly dependent at each period t.

The statistical implication of the adding up condition is as follows: post multiplying by $\varepsilon'_{.t}$ in equation (2.15) and taking expectations, we obtain

$$\begin{aligned} E\left[(\omega'\varepsilon_{.t})\varepsilon'_{.t}\right] &= 0, \\ \omega'\Sigma^* &= 0. \end{aligned}$$

This implies that the covariance of the disturbances is singular, that is, Σ^* is not invertible. Note that ω is the eigenvector of Σ^*. Thus, given the adding up condition (2.13) the Wald statistic (2.12) for testing the hypothesis (2.8) is not defined. One possible solution to this problem is to delete an equation. This conjecture is based on Barten (1969).

Barten considered the case where $\varepsilon_{.t}$ is i.i.d. $N(0, \Sigma^*)\ \forall\, t$. For this case he proved that the value of the likelihood function is invariant to the equation deleted assuming that the weights are equal and that $\iota' y_{.t} = 1 \, \forall \, t$, where ι is an $(n \times 1)$ vector of ones. This, in turn, implies that the value of the likelihood ratio (LR) statistic for testing (2.8) is invariant to the equation deleted.

We present a simple algebraic proof that the value of the Wald test statistic is invariant to the equation deleted. Our result is more general than Barten's: (i) we permit unequal weights, (ii) we let x_{1t} to be any explanatory variable instead of restricting to a constant, and (iii) we do not assume that the distribution of $\varepsilon_{.t}$ is normal. Indeed, if $\varepsilon_{.t}$ is $N\big(0, \Sigma^*\big)$ our proof implies that the value of the LR statistic is invariant to the equation deleted.

Dropping the nth equation, the SUR system becomes

$$y^{(n)} = Z^{(n)} \gamma^{(n)} + \varepsilon^{(n)}$$

where a superscript "n" on the vectors denotes that the nth equation is deleted and so the nth component is deleted from the corresponding vector, and $Z^{(n)}$ is equal to Z with the last row and column deleted.

The estimate of $\gamma^{(n)}$ and the variance of the estimate is given by

$$\begin{aligned} \hat{\gamma}^{(n)} &= (Z^{(n)\prime} Z^{(n)})^{-1} Z^{(n)\prime} y^{(n)}, \\ V(\hat{\gamma}^{(n)}) &= (Z^{(n)\prime} Z^{(n)})^{-1} Z^{(n)\prime} \Sigma^{(n)} Z^{(n)} (Z^{(n)\prime} Z^{(n)})^{-1}. \end{aligned}$$

Since $\Sigma^{(n)}$ is unknown, we replace the $\Sigma^{(n)}$ by a consistent estimator, $\hat{\Sigma}^{(n)}$, where

$$\begin{aligned} \hat{\Sigma}^{(n)} &= \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1,n-1} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \cdots & \hat{\sigma}_{2,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{n-1,1} & \hat{\sigma}_{n-1,2} & \cdots & \hat{\sigma}_{n-1,k} \end{bmatrix} \otimes I_T \\ &= \hat{\Sigma}^{*(n)} \otimes I_T, \end{aligned}$$

$\hat{\sigma}_{ij} = \frac{1}{T-k} \sum_{t=1}^{T} \hat{\varepsilon}_{it} \hat{\varepsilon}_{jt}, \; i, j = 1, 2, \ldots, n-1$, and I_T is an identity matrix of order T. Hence the estimated covariance matrix is

$$\hat{V}(\hat{\gamma}^{(n)}) = (Z^{(n)\prime} Z^{(n)})^{-1} Z^{(n)\prime} \hat{\Sigma}^{(n)} Z^{(n)} (Z^{(n)\prime} Z^{(n)})^{-1}.$$

Because of the property of least squares, the adding up restriction (2.13) also holds for the least squares residuals:

$$\hat{\varepsilon}_{1t} = -\frac{\omega_2}{\omega_1} \hat{\varepsilon}_{2t} - \frac{\omega_n}{\omega_1} \hat{\varepsilon}_{3t} - \cdots - \frac{\omega_n}{\omega_1} \hat{\varepsilon}_{nt} \quad \forall \, t. \tag{2.16}$$

Using (2.16), we obtain the following relation between $\hat{\varepsilon}_{.t}^{(n)}$ and $\hat{\varepsilon}_{.t}^{(1)}$:

(2.17)

$$\begin{bmatrix} \hat{\varepsilon}_{1t} \\ \hat{\varepsilon}_{2t} \\ \hat{\varepsilon}_{3t} \\ \vdots \\ \hat{\varepsilon}_{n-1,t} \end{bmatrix} = \begin{bmatrix} -\frac{\omega_2}{\omega_1} & -\frac{\omega_3}{\omega_1} & \dots & -\frac{\omega_{n-1}}{\omega_1} & -\frac{\omega_n}{\omega_1} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{\varepsilon}_{2t} \\ \hat{\varepsilon}_{3t} \\ \hat{\varepsilon}_{4t} \\ \vdots \\ \hat{\varepsilon}_{nt} \end{bmatrix} \quad \forall\, t.$$

This can be written compactly as

$$\hat{\varepsilon}_t^{(n)} = A\hat{\varepsilon}_t^{(1)} \tag{2.18}$$

where the matrix A is defined as

$$A = \begin{bmatrix} -\frac{\omega_2}{\omega_1} & -\frac{\omega_3}{\omega_1} & \dots & -\frac{\omega_{n-1}}{\omega_1} & -\frac{\omega_n}{\omega_1} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}. \tag{2.19}$$

Referring back to equations in (2.14) we obtain $\hat{\beta}_{.j}^{(n)} = A\hat{\beta}_{.j}^{(1)}$, $j = 2, 3, \dots, k$. In equation (2.18), postmultiplying by the transpose yields $(\hat{\varepsilon}_{.t}^{(n)}\hat{\varepsilon}_{.t}^{(n)\prime}) = A(\hat{\varepsilon}_{.t}^{(1)}\hat{\varepsilon}_{.t}^{(1)\prime})A'$. Since the matrix A is a constant, summing over t and dividing both sides by $(T-k)$ gives

$$\hat{\Sigma}^{*(n)} = A\hat{\Sigma}^{*(1)}A'. \tag{2.20}$$

Consider testing the null hypothesis $H_0^{(n)}$: $\beta_{.j}^{(n)} = 0$ or equivalently $H_0^{(n)}$: $r^{(n)}\gamma^{(n)} = 0$ where $r^{(n)} = I_{n-1} \otimes R_j'$. From (2.14), $\beta_{.j}^{(n)} = 0$ implies $\beta_{.j} = 0$ and hence H_0 (2.8). The Wald statistic for testing $H_0^{(n)}$ is

$$\begin{aligned} J_1^{(n)} &= \hat{\beta}_{.j}^{(n)\prime}\left[r^{(n)}\hat{V}(\hat{\gamma}^{(n)})r^{(n)\prime}\right]^{-1}\hat{\beta}_{.j}^{(n)} \\ &= \hat{\beta}_{.j}^{(n)\prime}\left[r^{(n)}(Z^{(n)\prime}Z^{(n)})^{-1}Z^{(n)\prime}(\Sigma^{*(n)} \otimes I_T)\right. \\ &\qquad \left. Z^{(n)}(Z^{(n)\prime}Z^{(n)})^{-1}r^{(n)\prime}\right]^{-1}\hat{\beta}_{.j}^{(n)}. \end{aligned} \tag{2.21}$$

Using the fact that $Z^{(n)} = Z^{(1)} = I_{n-1} \otimes X$ in (2.21) and simplifying, we obtain

$$\begin{aligned} J_1^{(n)} &= \hat{\beta}_{.j}^{(n)\prime}\left[\hat{\Sigma}^{*(n)} \otimes R_j'(X'X)^{-1}R_j\right]^{-1}\hat{\beta}_{.j}^{(n)} \\ &= \hat{\beta}_{.j}^{(n)\prime}\left[\hat{\Sigma}^{*(n)} \otimes c\right]^{-1}\hat{\beta}_{.j}^{(n)} \\ &= \hat{\beta}_{.j}^{(n)\prime}c^{-1}\left(\hat{\Sigma}^{*(n)}\right)^{-1}\hat{\beta}_{.j}^{(n)}. \end{aligned}$$

From (2.20) it follows that

$$\begin{aligned} J_1^{(n)} &= c^{-1}\hat{\beta}_{.j}^{(1)\prime}A'A'^{-1}\left(\hat{\Sigma}^{*(1)}\right)^{-1}A^{-1}A\hat{\beta}_{.j}^{(1)} \\ &= J_1^{(1)}. \end{aligned}$$

The result is summarized by the following theorem.

THEOREM 1: *Consider the SUR system defined by* (2.5) *and* (2.7). *Under the adding up restriction* (2.13), *the value of the Wald statistic for testing hypothesis* (2.8) *is invariant to the equation deleted, that is,* $J_1^{(i)} = J_1^{(1)},\ i = 2, 3, \ldots, n.$

If $\varepsilon_{.t}$ is i.i.d. $N\big(0, \Sigma^*\big)$ with Σ^* full rank, then we can use the likelihood ratio (LR) test to test the hypothesis (2.8). The likelihood function for (2.1) is

$$
\begin{aligned}
f(y_{.1}, y_{.2}, \ldots, y_{.T} | x_{.1}, x_{.2}, \ldots, x_{.T}) &= \prod_{t=1}^{T} p(y_{.t}|x_{.t}) \\
&= \prod_{t=1}^{T} (2\pi)^{\frac{n}{2}} |\Sigma^*|^{\frac{1}{2}} \times exp\left[-\frac{1}{2}\big(y_{.t} - \beta x_{.t}\big)' \Sigma^{*-1} \big(y_{.t} - \beta x_{.t}\big)\right]
\end{aligned}
\tag{2.22}
$$

where $\beta = \{\beta_{ij}\},\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, k$, is an $(n \times k)$ matrix of parameters.

The LR test is based on the logarithm of the likelihood ratio, which is the value of the constrained log-likelihood function minus the unconstrained log-likelihood function evaluated at the maximum likelihood estimators. Denoting $\mathcal{LR}$ as the log-likelihood ratio, we have

$$
\begin{aligned}
\mathcal{LR} &= L^0 - L \\
&= -\frac{T}{2}\Big[\log|\hat{\Sigma}^{*0}| - \log|\hat{\Sigma}^*|\Big],
\end{aligned}
\tag{2.23}
$$

where L^0 represents the constrained likelihood function and where $\hat{\Sigma}^{*0}$ and $\hat{\Sigma}^*$ are the constrained and unconstrained maximum likelihood estimators of the covariance matrices respectively.

The LR test statistic is -2 times the logarithm of the likelihood ratio:

$$
\begin{aligned}
J_2 &= -2\mathcal{LR} \\
&= T\big[\log|\hat{\Sigma}^{**}| - \log|\hat{\Sigma}^*|\big].
\end{aligned}
\tag{2.24}
$$

Under the null hypothesis, J_2 is asymptotically distributed as a chi-square with n degrees of freedom.

To prove that the value of the LR statistic is invariant to the equation deleted we use the fact that J_1 is a monotonic transformation of J_2. The monotonic transformation is given by

$$
J_1 = \frac{\big(T - n - 1\big)}{n} \left(exp\left[\frac{J_2}{T}\right] - 1\right). \tag{2.25}
$$

Since $J_1^{(i)} = J_1^{(1)},\ i = 2, 3, \ldots, n$, it follows from (2.25) that $J_2^{(i)} = J_2^{(1)},\ i = 2, 3, \ldots, n$. Thus the likelihood ratio test statistic for testing H_0 is also invariant to the equation deleted.

Corollary 1: *Consider the SUR system defined by* (2.5) *and* (2.7) *with* $\varepsilon_{.t}$ *being i.i.d.* $N\big(0, \Sigma^*\big)$. *Under the adding up restriction* (2.13), *the value of the LR statistic for testing hypothesis* (2.8) *is invariant to the equation deleted, that is,* $J_2^{(i)} = J_2^{(1)}$, $i = 2, 3, \ldots, n$.

3. Sur System with Heteroskedasticity

Consider the system of equations in (2.1) where the $\varepsilon_{.t}$'s are independent with zero mean and the covariance matrix is time varying:

$$\left.\begin{array}{lcl} E(\varepsilon_{.t}) & = & 0, \\ E(\varepsilon_{.t}\varepsilon_{.t}^{'}) & = & \Sigma_t^* \end{array}\right\} \quad \forall\, t \tag{3.1}$$

where $\Sigma_t^* = \{\sigma_{t,ij}\}$, $i, j = 1, 2, \ldots, n$, is an $(n \times n)$ time varying contemporaneous covariance matrix.

Here we assume that y_{it}'s are independent across time, but have time varying variances. In any period t, there is a non-zero correlation between contemporaneous disturbances in the ith and jth equations, but zero correlation between all lagged disturbances. The covariance matrix for the SUR model in this case, Σ, is $\{\Sigma_{ij}\}$, $i, j = 1, 2, \ldots, n$, where Σ_{ij} is the ijth block defined by

$$\Sigma_{ij} = \begin{bmatrix} \sigma_{1,ij} & 0 & \ldots & 0 \\ 0 & \sigma_{2,ij} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_{T,ij} \end{bmatrix}. \tag{3.2}$$

Note that $\sigma_{t,ij}$ denotes the contemporaneous covariance between the ith and jth equation at period t. Thus, in the heteroskedastic case Σ_{ij} in (2.6) is replaced by (3.2).

The null hypothesis to be tested is again (2.8). Following White (1980), we use the heteroskedasticity consistent (HC) covariance matrix estimator in the Wald statistic to test this hypothesis. The advantage of this approach is that we do not have to specify the functional form of heteroskedasticity, which is usually unknown. In this approach the unknown Σ_{ij} is replaced by $\{\hat{\Sigma}_{ij}\}$, $i, j = 1, 2, \ldots, n$, where

$$\hat{\Sigma}_{ij} = \begin{bmatrix} \hat{\varepsilon}_{i1}\hat{\varepsilon}_{j1} & 0 & \ldots & 0 \\ 0 & \hat{\varepsilon}_{i2}\hat{\varepsilon}_{j2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \hat{\varepsilon}_{iT}\hat{\varepsilon}_{jT} \end{bmatrix}. \tag{3.3}$$

See also Hamilton (p. 218, (1994)).

Define

$$\eta_t = \varepsilon_{.t} \otimes x_{.t}, \tag{3.4}$$

where η_t is an $(nk \times 1)$ column vector and $x_{.t} = (x_{1t}, x_{2t}, \ldots, x_{kt})'$, an $(k \times 1)$ vector of explanatory variables at period t. Post-multiplying both sides of equation (3.4) by its transpose we obtain

$$\eta_t \eta_t' = \left(\varepsilon_{.t} \otimes x_{.t}\right)\left(\varepsilon_{.t} \otimes x_{.t}\right)' = \varepsilon_{.t}\varepsilon_{.t}' \otimes x_{.t}x_{.t}'. \tag{3.5}$$

The estimator of $\eta_t\eta_t'$ is obtained by replacing the population variables by the corresponding least-squares estimates:

$$\hat{\eta}_t\hat{\eta}_t' = \hat{\varepsilon}_{.t}\hat{\varepsilon}_{.t}' \otimes x_{.t}x_{.t}' = \hat{\Sigma}_t^* \otimes x_{.t}x_{.t}'. \tag{3.6}$$

Summing both sides of equation (3.6) over t, we obtain

$$\sum_{t=1}^{T} \hat{\eta}_t\hat{\eta}_t' = \sum_{t=1}^{T}\left(\hat{\varepsilon}_{.t}\hat{\varepsilon}_{.t}' \otimes x_{.t}x_{.t}'\right)$$
$$= \begin{bmatrix} \sum_{t=1}^{T} x_{.t}x_{.t}'\hat{\varepsilon}_{1t}^2 & \sum_{t=1}^{T} x_{.t}x_{.t}'\hat{\varepsilon}_{1t}\hat{\varepsilon}_{2t} & \ldots & \sum_{t=1}^{T} x_{.t}x_{.t}'\hat{\varepsilon}_{1t}\hat{\varepsilon}_{nt} \\ \sum_{t=1}^{T} x_{.t}x_{.t}'\hat{\varepsilon}_{2t}\hat{\varepsilon}_{1t} & \sum_{t=1}^{T} x_{.t}x_{.t}'\hat{\varepsilon}_{2t}^2 & \ldots & \sum_{t=1}^{T} x_{.t}x_{.t}'\hat{\varepsilon}_{2t}\hat{\varepsilon}_{nt} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^{T} x_{.t}x_{.t}'\hat{\varepsilon}_{nt}\hat{\varepsilon}_{1t} & \sum_{t=1}^{T} x_{.t}x_{.t}'\hat{\varepsilon}_{nt}\hat{\varepsilon}_{2t} & \ldots & \sum_{t=1}^{T} x_{.t}x_{.t}'\hat{\varepsilon}_{nt}^2 \end{bmatrix}.$$

Recall that $\sum_{t=1}^{T} x_{.t}x_{.t}' = X'X$, and hence, $\sum_{t=1}^{T} x_{.t}x_{.t}'\hat{\varepsilon}_{it}\hat{\varepsilon}_{jt} = X'\hat{\Sigma}_{ij}X$, $i, j = 1, 2, \ldots, n$. Thus, the above expression can be written as

$$\sum_{t=1}^{T} \hat{\eta}_t\hat{\eta}_t' = \begin{bmatrix} X'\hat{\Sigma}_{11}X & X'\hat{\Sigma}_{12}X & \ldots & X'\hat{\Sigma}_{1n}X \\ X'\hat{\Sigma}_{21}X & X'\hat{\Sigma}_{22}X & \ldots & X'\hat{\Sigma}_{2n}X \\ \vdots & \vdots & \ddots & \vdots \\ X'\hat{\Sigma}_{n1}X & X'\hat{\Sigma}_{n2}X & \ldots & X'\hat{\Sigma}_{nn}X \end{bmatrix} = Z'\hat{\Sigma}Z. \tag{3.7}$$

Substituting $Z = I_n \otimes X$, $Z'\hat{\Sigma}Z = \sum_{t=1}^{T} \hat{\eta}_t\hat{\eta}_t'$ and using (3.6), the HC estimator of the variance of $\hat{\gamma}$ is

$$\begin{aligned} \hat{V}(\hat{\gamma}) &= \left(I_n \otimes (X'X)^{-1}\right)\left(\sum_{t=1}^{T} \hat{\eta}_t\hat{\eta}_t'\right)\left(I_n \otimes (X'X)^{-1}\right) \\ &= \sum_{t=1}^{T}\left(I_n \otimes (X'X)^{-1}\right)\left(\hat{\Sigma}_t^* \otimes x_{.t}x_{.t}'\right)\left(I_n \otimes (X'X)^{-1}\right) \\ &= \sum_{t=1}^{T}\left(\hat{\Sigma}_t^* \otimes (X'X)^{-1}x_{.t}x_{.t}'(X'X)^{-1}\right). \end{aligned}$$

Thus,

$$\hat{V}\left(\hat{\gamma}\right) = \sum_{t=1}^{T}\left(\hat{\Sigma}_t^* \otimes Q_t\right) \tag{3.8}$$

where $Q_t = \left(X^{'}X\right)^{-1} x_{.t}x_{.t}^{'}\left(X^{'}X\right)^{-1}$ is an $(k \times k)$ matrix which depends on t.

The Wald statistic for testing the hypothesis (2.8) is then given by

$$J_3 = \hat{\beta}_{.j}^{'}\left[r\hat{V}(\hat{\gamma})r^{'}\right]^{-1}\hat{\beta}_{.j}. \tag{3.9}$$

Substituting $r = I_n \otimes R_j^{'}$ and the expression for $\hat{V}\left(\hat{\gamma}\right)$, the Wald statistic can be written as

$$\begin{aligned} J_3 &= \hat{\beta}_{.j}^{'}\left[\left(I_n \otimes R_j^{'}\right)\left(\sum_{t=1}^{T}(\hat{\Sigma}_t^* \otimes Q_t)\right)\left(I_n \otimes R_j\right)\right]^{-1}\hat{\beta}_{.j} \\ &= \hat{\beta}_{.j}^{'}\left[\sum_{t=1}^{T}\left(\hat{\Sigma}_t^* \otimes R_j^{'}Q_tR_j\right)\right]^{-1}\hat{\beta}_{.j} \\ &= \hat{\beta}_{.j}^{'}\left[\sum_{t=1}^{T}\left(\hat{\Sigma}_t^* \otimes q_t\right)\right]^{-1}\hat{\beta}_{.j} \end{aligned}$$

where q_t is the jth diagonal element of Q_t. Hence,

$$J_3 = \hat{\beta}_{.j}^{'}\left[\sum_{t=1}^{T} q_t\hat{\Sigma}_t^*\right]^{-1}\hat{\beta}_{.j}. \tag{3.10}$$

Under the null hypothesis, J_3 has an asymptotic chi-square distribution with n degrees of freedom under some regularity assumptions. See for example, Hamilton (Assumption 8.6, (1994)).

Adding Up

Suppose we now impose the adding up restriction (2.13). The adding up condition implies the following restriction on the covariance matrices:

$$\omega^{'}\Sigma_t^* = 0 \quad \forall\, t. \tag{3.11}$$

Hence, the covariance matrix Σ_t^* is singular for all t. Because of the property of least squares, the above condition (3.11) also holds for the corresponding samples; that is, $\omega^{'}\hat{\Sigma}_t^* = 0 \ \forall\ t$. Since ω is a non-zero constant, the adding up condition would imply $\omega^{'}\sum_{t=1}^{T} q_t\hat{\Sigma}_t^* = 0$ for some non-zero scalars q_t such that not all q_t's are equal to zero. The implication of the above condition is that the weighted sum of the covariance matrices in (3.10) is not invertible, and thus, the Wald statistic (3.10) for testing the hypothesis (2.8) is not defined. Again we resort to the same technique of dropping an equation as a solution to the problem. We present an algebraic proof that the value of the HC Wald test statistic is invariant to the equation deleted.

The adding up restriction implies (from equation (2.18)) that

$$\hat{\varepsilon}^{(n)}_{.t} = A\hat{\varepsilon}^{(1)}_{.t}.$$

Multiplying the above equation by its transpose we get

$$\begin{aligned} \hat{\varepsilon}^{(n)}_{.t}\hat{\varepsilon}^{(n)\prime}_{.t} &= A\hat{\varepsilon}^{(1)}_{.t}\hat{\varepsilon}^{(1)\prime}_{.t}A', \\ \hat{\Sigma}^{*(n)}_{t} &= A\hat{\Sigma}^{*(1)}_{t}A'. \end{aligned} \tag{3.12}$$

The HC Wald statistic (with the nth equation dropped) is given by

$$\begin{aligned} J_3^{(n)} &= \hat{\beta}^{(n)\prime}_{.j}\left[r^{(n)}\hat{V}\big(\hat{\gamma}^{(n)}\big)r^{(n)\prime}\right]^{-1}\hat{\beta}^{(n)}_{.j} \\ &= \hat{\beta}^{(n)'}_{.j}\left[r^{(n)}\big(Z^{(n)\prime}Z^{(n)}\big)^{-1}\big(\hat{\Sigma}^{*(n)}_t\otimes x_{.t}x'_{.t}\big)\big(Z^{(n)\prime}Z^{(n)}\big)^{-1}r^{(n)\prime}\right]^{-1}\hat{\beta}^{(n)}_{.j}. \end{aligned}$$

Substituting $r = I_{n-1}\otimes R'_j$ and $Z^{(n)} = Z^{(1)} = I_{n-1}\otimes X$ in the above expression and simplifying, we obtain

$$\begin{aligned} J_3^{(n)} &= \hat{\beta}^{(n)\prime}_{.j}\left[\big(I_{n-1}\otimes R'_j\big)\Big(\sum_{t=1}^{T}\big(\hat{\Sigma}^{*(n)}_t\otimes Q_t\big)\Big)\big(I_{n-1}\otimes R_j\big)\right]^{-1}\hat{\beta}^{(n)}_{.j} \\ &= \hat{\beta}^{(n)\prime}_{.j}\left[\sum_{t=1}^{T}\big(\hat{\Sigma}^{*(n)}\otimes R'_jQ_tR_j\big)\right]^{-1}\hat{\beta}^{(n)}_{.j} \\ &= \hat{\beta}^{(n)\prime}_{.j}\left[\sum_{t=1}^{T}\big(\hat{\Sigma}^{*(n)}_t\otimes q_t\big)\right]^{-1}\hat{\beta}^{(n)}_{.j} \\ &= \hat{\beta}^{(n)\prime}_{.j}\left[\sum_{t=1}^{T}q_t\hat{\Sigma}^{*(n)}_t\right]^{-1}\hat{\beta}^{(n)}_{.j} \\ &= \hat{\beta}^{(n)\prime}_{.j}\left[\sum_{t=1}^{T}q_t\big(A\hat{\Sigma}^{*(1)}_tA'\big)\right]^{-1}\hat{\beta}^{(n)}_{.j} \\ &= \hat{\beta}^{(1)\prime}_{.j}A'\left[A\Big[\sum_{t=1}^{T}q_t\hat{\Sigma}^{*(1)}_t\Big]A'\right]^{-1}A\hat{\beta}^{(1)}_{.j} \\ &= \hat{\beta}^{(1)\prime}_{.j}A'A'^{-1}\left[\sum_{t=1}^{T}q_t\hat{\Sigma}^{*(1)}_t\right]^{-1}A^{-1}A\hat{\beta}^{(1)}_{.j} \\ &= J_3^{(1)}. \end{aligned}$$

This proves the following theorem on the invariance property for the heteroskedastic case.

THEOREM 2: *Consider a SUR system defined by* (2.5) *and* (3.2). *Under the adding up restriction* (2.13), *the value of the HC Wald statistic for testing the hypothesis* (2.8) *is invariant to the equation deleted, that is,* $J_3^{(i)} = J_3^{(1)}$, $i = 2, 3, \ldots, n$.

4. SUR System with Heteroskedasticity and Autocorrelation

Consider the system of equations in (2.1) where the $\varepsilon_{.t}$'s are serially correlated up to a order of lag one and have time varying covariance matrices:

$$E(\varepsilon_{.t}) = 0 \quad \forall\, t,$$

$$E(\varepsilon_{.t}\varepsilon_{.s}^{'}) = \begin{cases} \Sigma_{t,t}^{*} & \text{if } t = s, \\ \Sigma_{t,t-1}^{*} & \text{if } t - s = 1, \\ \Sigma_{t-1,t}^{*} & \text{if } s - t = 1, \\ 0 & \text{if } |t-s| \geq 2 \end{cases} \tag{4.1}$$

where $\Sigma_{t,t}^{*} = \{\sigma_{tt,ij}\}$, $i, j = 1, 2, \ldots,\ n$, is a $(n \times n)$ time dependent contemporaneous covariance matrix and $\Sigma_{t,t-1}^{*} = \{\sigma_{tt-1,ij}\}$, $i, j = 1, 2,\ \ldots, n$, is a $(n \times n)$ time dependent lag-1 covariance matrix.

This is the general setup where the disturbances have variances which vary with time and are serially correlated. The above setup gives a simple auto-correlation structure to the disturbances: in any period t the errors are serially correlated with each other up to a maximum lag of length one. We have assumed a lag of length one purely for expositional convenience. Our method is applicable for any number of finite lags. The covariance matrix for the SUR model in this case, Σ, is $\{\Sigma_{ij}\}$, $i, j, 1, 2, \ldots, n$, where Σ_{ij} is the ijth block is defined as

$$\Sigma_{ij} = \begin{bmatrix} \sigma_{11,ij} & \sigma_{12,ij} & 0 & 0 & \ldots & 0 & 0 \\ \sigma_{21,ij} & \sigma_{22,ij} & \sigma_{23,ij} & 0 & \ldots & 0 & 0 \\ 0 & \sigma_{32,ij} & \sigma_{33,ij} & \sigma_{34,ij} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \sigma_{T,T-1,ij} & \sigma_{TT,ij} \end{bmatrix}. \tag{4.2}$$

Note that $\sigma_{tt,ij}$ denotes the contemporaneous covariance between the ith and jth equation at period t and $\sigma_{tt-1,ij}$ denotes the one period lagged covariance between the ith and jth equation across periods t and $t-1$. Hence, in the case with heteroskedasticity and autocorrelation, Σ_{ij} in equation (2.6) is replaced by (4.2). The hypothesis to be tested, (2.8), remains the same. We shall use the Wald statistic based on the heteroskedaticity and autocorrelation consistent (HAC) covariance matrix to test this hypothesis. The unknown Σ_{ij} is replaced by $\hat{\Sigma}_{ij}$

where

$$(4.3)\quad \hat{\Sigma}_{ij} = \begin{bmatrix} \hat{\varepsilon}_{1i}\hat{\varepsilon}_{1j} & \hat{\varepsilon}_{1i}\hat{\varepsilon}_{2j} & 0 & 0 & \dots & 0 & 0 \\ \hat{\varepsilon}_{2i}\hat{\varepsilon}_{1j} & \hat{\varepsilon}_{2i}\hat{\varepsilon}_{2j} & \hat{\varepsilon}_{2i}\hat{\varepsilon}_{3j} & 0 & \dots & 0 & 0 \\ 0 & \hat{\varepsilon}_{3i}\hat{\varepsilon}_{2j} & \hat{\varepsilon}_{3i}\hat{\varepsilon}_{3j} & \hat{\varepsilon}_{3i}\hat{\varepsilon}_{4j} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \hat{\varepsilon}_{ni}\hat{\varepsilon}_{T-1,j} & \hat{\varepsilon}_{Ti}\hat{\varepsilon}_{Tj} \end{bmatrix}$$

in the HAC Wald statistic.

The HAC estimator of variance of $\hat{\gamma}$ is given by

$$\hat{V}(\hat{\gamma}) = \left(Z^{'}Z\right)^{-1}Z^{'}\hat{\Sigma}Z\left(Z^{'}Z\right)^{-1}$$

where $\hat{\Sigma}$ is defined above. From the definition of $\hat{\eta}_t$ in (3.6) it follows that

$$(4.4)\qquad Z^{'}\hat{\Sigma}Z = \sum_{t=1}^{T}\hat{\eta}_t\hat{\eta}_t^{'} + \sum_{\nu=1}^{p}\Big(1 - \frac{\nu}{p+1}\Big)\sum_{t=\nu+1}^{T}\Big[\hat{\eta}_t\hat{\eta}_{t-\nu}^{'} + \hat{\eta}_{t-\nu}\hat{\eta}_t^{'}\Big]$$

where the weights are the same as in Newey and West (1987), and p is a lag length beyond which we are willing to assume that the correlation between η_t and $\eta_{t-\nu}$ is essentially equal to zero. See Hamilton (p. 317, 1994). In our case the maximum lag length is one, that is, $p = 1$. Thus, (4.4) reduces to

$$(4.5)\qquad Z^{'}\hat{\Sigma}Z = \sum_{t=1}^{T}\hat{\eta}_t\hat{\eta}_t^{'} + \frac{1}{2}\sum_{t=2}^{T}\hat{\eta}_t\hat{\eta}_{t-1}^{'} + \frac{1}{2}\sum_{t=2}^{T}\hat{\eta}_{t-1}\hat{\eta}_t^{'}.$$

Defining $\hat{\eta}_t\hat{\eta}_t^{'} = \hat{\Sigma}^*_{t,t} \otimes x_{.t}x^{'}_{.t}$, $\hat{\eta}_t\hat{\eta}_{t-1}^{'} = \hat{\Sigma}^*_{t,t-1} \otimes x_t x^{'}_{.t-1}$ and $\hat{\eta}_{t-1}\hat{\eta}_t^{'} = \hat{\Sigma}^*_{t-1,t} \otimes x_{.t-1}x^{'}_{.t}$ in (4.5) and substituting, we obtain

$$\begin{aligned} Z^{'}\hat{\Sigma}Z = &\sum_{t=1}^{T}\big(\hat{\Sigma}^*_{t,t} \otimes x_{.t}x^{'}_{.t}\big) + \frac{1}{2}\sum_{t=2}^{T}\big(\hat{\Sigma}^*_{t,t-1} \otimes x_{.t}x^{'}_{.t-1}\big) \\ &+ \frac{1}{2}\sum_{t=2}^{T}\big(\hat{\Sigma}^*_{t-1,t} \otimes x_{.t-1}x^{'}_{.t}\big). \end{aligned}$$

Hence, the HAC estimator of the variance of $\hat{\gamma}$ is given by

$$\hat{V}\left(\hat{\gamma}\right) = \left(I_n \otimes (X^{'}X)^{-1}\right)\left[Z^{'}\hat{\Sigma}Z\right]\left(I_n \otimes (X^{'}X)^{-1}\right).$$

Substituting the expression for $Z^{'}\hat{\Sigma}Z$, we obtain

$$\begin{aligned} \hat{V}\left(\hat{\gamma}\right) = &\sum_{t=1}^{T}\Big[\hat{\Sigma}^*_{t,t} \otimes (X^{'}X)^{-1}x_{.t}x^{'}_{.t}(X^{'}X)^{-1}\Big] + \frac{1}{2}\sum_{t=2}^{T}\Big[\hat{\Sigma}^*_{t,t-1} \otimes \\ &(X^{'}X)^{-1}x_{.t}x^{'}_{.t-1}(X^{'}X)^{-1} + \hat{\Sigma}^*_{t-1,t} \otimes (X^{'}X)^{-1}x_{.t-1}x^{'}_{.t}(X^{'}X)^{-1}\Big]. \end{aligned}$$

Define the following $(k \times k)$ matrix:

$$Q_{t,s} = \left(X'X\right)^{-1} x_{.t} x'_{.s} \left(X'X\right)^{-1}. \tag{4.6}$$

Finally, using (4.6)

$$\begin{aligned} \hat{V}(\hat{\gamma}) = \sum_{t=1}^{T} \left(\hat{\Sigma}^*_{t,t} \otimes Q_{t,t}\right) + \frac{1}{2} \sum_{t=2}^{T} \Big[\left(\hat{\Sigma}^*_{t,t-1} \otimes Q_{t,t-1}\right) \\ + \left(\hat{\Sigma}^*_{t-1,t} \otimes Q_{t-1,t}\right) \Big]. \end{aligned} \tag{4.7}$$

The HAC Wald statistic for testing the hypotheis (2.8) is given by

$$J_4 = \hat{\beta}'_{.j} \left[r \hat{V}(\hat{\gamma}) r' \right]^{-1} \hat{\beta}_{.j}. \tag{4.8}$$

Substituting $r = I_n \otimes R'_j$ and the expression for variance (4.7), the Wald statistic can be written as

$$\begin{aligned} J_4 &= \hat{\beta}'_{.j} \Bigg[\sum_{t=1}^{T} \left(\hat{\Sigma}^*_{t,t} \otimes R'_j Q_{t,t} R_j\right) + \frac{1}{2} \sum_{t=2}^{T} \Big[\left(\hat{\Sigma}^*_{t,t-1} \otimes R'_j Q_{t,t-1} R_j\right) \\ &\quad + \left(\hat{\Sigma}^*_{t-1,t} \otimes R'_j Q_{t-1,t} R_j\right) \Big] \Bigg]^{-1} \hat{\beta}_{.j} \\ &= \hat{\beta}'_{.j} \Bigg[\sum_{t=1}^{T} \left(\hat{\Sigma}^*_{t,t} \otimes q_{t,t}\right) + \frac{1}{2} \sum_{t=2}^{T} \Big[\left(\hat{\Sigma}^*_{t,t-1} \otimes q_{t,t-1}\right) \\ &\quad + \left(\hat{\Sigma}^*_{t-1,t} \otimes q_{t-1,t}\right) \Big] \Bigg]^{-1} \hat{\beta}_{.j} \end{aligned}$$

where $q_{t,s}$ is the jth diagonal element of $Q_{t,s}$ in (4.6). Hence, the Wald statistic can be re-written as

$$J_4 = \hat{\beta}'_{.j} \left[\sum_{t=1}^{T} q_{t,t} \hat{\Sigma}^*_{t,t} + \frac{1}{2} \sum_{t=2}^{T} \left[q_{t,t-1} \hat{\Sigma}^*_{t,t-1} + q_{t-1,t} \hat{\Sigma}^*_{t-1,t} \right] \right]^{-1} \hat{\beta}_{.j}. \tag{4.9}$$

Under the null hypothesis, J_4 will have an asymptotic chi-square distribution with n degrees of freedom under some regularity conditions. See, for example, Hamilton (p. 225, (1994)).

Adding Up

Suppose now we impose the adding up condition (2.13). The adding up condition would imply the following restriction on the covariance matrices:

$$\omega' \Sigma^*_{t,s} = 0, \quad t, s = 1, 2, \ldots, T \text{ and } |t - s| \leq 1. \tag{4.10}$$

Equation (4.10) implies that the covariance matrix $\Sigma^*_{t,s}$ is singular for $t, s = 1, 2, \ldots, T$, and $|t - s| \leq 1$. The least squares property implies that the above condition is also true in the sample: $\omega' \Sigma^*_{t,s} =$

0, $t, s = 1, 2, \ldots, n,$ and $|t-s| \le 1$. Since ω is a non-zero weight vector the adding up condition implies $\omega' \sum_{t=1}^{T} q_{t,t}\hat{\Sigma}^*_{t,t} = 0$ for some scalars $q_{t,t}$ such that not all $q_{t,t}$'s are zeros. Condition (4.10) also implies $\omega' \sum_{t=2}^{T} q_{t,t-1}\hat{\Sigma}^*_{t,t-1} = 0$ for some non-zero scalars $q_{t,t-1}$ and similarly $\omega' \sum_{t=2}^{T} q_{t-1,t}\hat{\Sigma}^*_{t-1,t} = 0$ for some non-zero scalars $q_{t-1,t}$. Thus, the HAC Wald statistic in (4.9) is not defined because the expression in parenthesis in (4.9) is not invertible. Again, we resort to the same technique of dropping an equation as a solution to the problem. We present an algebraic proof that the value of the HAC Wald statistic is invariant to the equation dropped.

From equation (2.18) we have

$$\hat{\varepsilon}^{(n)}_{.t} = A\hat{\varepsilon}^{(1)}_{.t}.$$

Multiplying the above expression by its transpose at period s, we obtain

$$\begin{aligned} \hat{\varepsilon}^{(n)}_{.t}\hat{\varepsilon}^{(n)'}_{.s} &= A\hat{\varepsilon}_{.t}\hat{\varepsilon}^{(1)'}_{.s}A', \\ \hat{\Sigma}^{*(n)}_{t,s} &= A\hat{\Sigma}^{*(1)}_{t,s}A', \end{aligned} \tag{4.11}$$

$t, s = 1, 2, \ldots, n,$ and $|t-s| \le 1$. The Wald statistic in (4.9) with the nth equation dropped is

$$J^{(n)}_4 = \hat{\beta}^{(n)'}_{.j}\left[\sum_{t=1}^{T} q_{t,t}\hat{\Sigma}^{*(n)}_{t,t} + \frac{1}{2}\sum_{t=2}^{T}\Big[q_{t,t-1}\hat{\Sigma}^{*(n)}_{t,t-1} + q_{t-1,t}\hat{\Sigma}^{*(n)}_{t-1,t}\Big]\right]^{-1}\hat{\beta}^{(n)}_{.j}.$$

Using the transformation in (4.11) and substituting $\hat{\beta}^{(n)}_{.j} = A\hat{\beta}^{(1)}_{.j}$, we obtain

$$\begin{aligned} J^{(n)}_4 = \hat{\beta}^{(1)'}_{.j}A'\Bigg[\sum_{t=1}^{T} q_{t,t}\big(A\hat{\Sigma}^{*(1)}_{t,t}A'\big) + \frac{1}{2}\sum_{t=2}^{T}\Big[q_{t,t-1}\big(A\hat{\Sigma}^{*(1)}_{t,t-1}A'\big) + \\ q_{t-1,t}\big(A\hat{\Sigma}^{*(1)}_{t-1,t}A'\big)\Big]\Bigg]^{-1}A\hat{\beta}^{(1)}_{.j} \end{aligned}$$

$$\begin{aligned} = \hat{\beta}^{(1)'}_{.j}A'\Bigg[A\Bigg(\sum_{t=1}^{T} q_{t,t}\hat{\Sigma}^{*(1)}_{t,t} + \frac{1}{2}\sum_{t=2}^{T}\Big[q_{t,t-1}\hat{\Sigma}^{*(1)}_{t,t-1} + \\ q_{t-1,t}\hat{\Sigma}^{*(1)}_{t-1,t}\Big]\Bigg)A'\Bigg]^{-1}A\hat{\beta}^{(1)}_{.j} \end{aligned}$$

$$= \hat{\beta}^{(1)'}_{.j} A' A'^{-1} \left[\sum_{t=1}^{T} q_{t,t} \hat{\Sigma}^{*(1)}_{t,t} + \frac{1}{2} \sum_{t=2}^{T} \left[q_{t,t-1} \hat{\Sigma}^{*(1)}_{t,t-1} + q_{t-1,t} \hat{\Sigma}^{*(1)}_{t-1,t} \right] \right]^{-1} A^{-1} A \hat{\beta}^{(1)}_{.j}$$

$$= J_4^{(1)}.$$

This proves the following theorem on the invariance property for the heteroskedasticity and autocorrelation case.

THEOREM 3: *Consider the SUR system defined by* (2.5) *and* (4.2). *Under the adding up restriction* (2.13), *the value of the HAC Wald statistic for testing hypothesis* (2.8) *is invariant to the equation deleted, that is,* $J_4^{(i)} = J_4^{(1)}$, $i = 2, 3, \ldots, n$.

5. Example

We consider the Sharpe-Lintner version of the Capital Asset Pricing Model (CAPM), presented in Campbell, Lo and MacKinlay (1997), to illustrate the testing methodology. Assume that investors can borrow and lend at a risk free rate of return. Define $Z_{.t}$ as an $(n \times 1)$ vector of excess returns for n assets. For these n assets, the excess returns are described by

$$Z_{.t} = \alpha + \beta Z_{mt} + \varepsilon_{.t} \tag{5.1}$$

where α and β are $(n \times 1)$ parameter vectors, Z_{mt} is the period t excess return on the market portfolio and $\varepsilon_{.t}$ is an $(n \times 1)$ vector of disturbances.

In the CAPM, the return on the market portfolio is a weighted sum of returns on the individual assets:

$$\omega' Z_{.t} = Z_{mt} \quad \forall\, t; \tag{5.2}$$

for example, see Jagannathan and McGrattan (1995). The above condition implies that the covariance matrix of disturbances in model (5.1) is singular.

We consider testing the null hypothesis $H_0\colon \alpha = 0$. The rejection of H_0 is usually interpreted as the rejection of CAPM. We test H_0 using three test statistics: J_1 from (2.12), J_3 from (3.10) and J_4 from (4.9). The tests are conducted using a thirty-year sample for CRSP monthly returns on ten size sorted portfolios ($n = 10$). The ten portfolios include all stocks listed on the New York and on the American Stock exchanges. The one-month US Treasury bill is used as the risk-free asset. The sample extends from January 1965 through December 1994. Tests are conducted for the overall period, three ten-year subperiods and six five-year subperiods.

We calculate the test statistics J_1, J_3 and J_4 in two sets of experiments. In the spirit of Campbell et. al. (1997), in the first set of experiments, the statistics are calculated for the full system (5.1) using the CRSP value-weighted return (VWRETD) as the market portfolio. This market portfolio does not satisfy the adding up condition (5.2). In the second set of experiments the test statistics are calculated with one equation deleted from (5.1) and a market portfolio satisfying (5.2). Note that in the first set of experiments, the singularity problem is circumvented by using the VWRETD index as a proxy for the market portfolio instead of using (5.2) as suggested by the CAPM.

In practice, due to re-balancing, the portfolio weights in (5.2) may change from period to period. In the second set of experiments, the test statistics in each sample subperiod are calculated by fixing the portfolio weights at mean of the sample subperiod.

The empirical results for the Classical SUR model are reported in Table 1 and the results for SUR model with heteroskedasticity and autocorrelation are reported in Table 2.

An inspection of Table 1 shows the values of the test statistics in panel A and B look similar for the five-year subsamples; the exception is 1/80-12/84. For the ten-year subsamples, the values are similar only for the 1/65-12/74 period. Finally, the values differ substantially for the thirty-year subsamples.

Focusing on the values of the test statistics can be misleading when comparing panels A and B. In panel A, the J_1 statistic is distributed asymtotically as chi-square with 10 degrees of freedom whereas it has 9 degrees of freedom in panel B. For this reason, it is more useful to compare p-values across the two panels. At the 5% significance level, the null hypothesis is rejected for 3 five-year periods in panel A and 2 five-year periods in panel B; the null is also rejected for 1 ten-year period in panel A. Note that the p-values tend to be substantially higher in panel B.

In Table 2, the test statistics and, hence, the p-values for the heteroskedasticity case (lag-0) look similar to those in Table 1. This may be evidence against heteroskedasticity. Again, using the HC Wald test, the null hypothesis is rejected for 3 five-year periods in panel A and 2 five-year periods in panel B; the null is rejected for 1 ten-year period and the thirty-year period in panel A.

For the HAC Wald tests (lag-1 and lag-4), the values of the test statistics are typically larger than those for the HC or i.i.d. case. Futhermore, the higher the lag the larger the value of the test statistic. As a consequence, the null is rejected for all subsamples in panel A when lag equals 4; in panel B, the null is rejected in 7 out of 10 subsamples.

6. Concluding Comments

We have established the invariance of the robust Wald statistic to the equation deleted in SUR systems with adding up restrictions. We introduced our method of proof using the classical SUR system. We then proved the results for the SUR systems with heteroskedasticity and autocorrelation of unknown form. The latter cases are clearly extensions of Barten (1969). Furthermore, the adding up restriction used in this paper is also more general than the one considered by Barten. Our proofs follow a straightforward algebraic approach. As a consequence, unlike the previous literature, our proofs do not depend on any distributional assumptions except the existence of first and second moments.

Department of Economics, University of Iowa, 108 Pappajohn Business Admn. Bldg., Iowa City, IA 52242-1000.

References

Barten, A.P. (1969): "Maximum likelihood estimation of a complete system of demand equations," *European Economic Review*, 17-73.

Berndt, E.R., and N.E. Savin (1975): "Estimation and hypothesis testing in singular equation systems with autoregressive disturbances," *Econometrica*, 43, 937-957.

Campbell, J. Y., A. W. Lo and A. C. MacKinlay (1997): *The Econometrics of Financial Markets,* Princeton University Press, New Jersey.

Greene W.H. (1997): *Econometric Analysis,* Prentice Hall, New Jersey.

Hamilton, J.D. (1994): *Time Series Analysis,* Princeton University Press, New Jersey.

Jagannathan, R. and E.R. McGrattan (1995): "The CAPM Debate," *Federal Reserve Bank of Minneapolis, Quarterly Review* Fall 1995.

Mandy, D.M. and C. Martins-Filho (1993): "SUR under additive Heteroskedasticity: Theory and Share Equation Applications," *Journal of Econometrics*, 58, 315-346.

McGuire, T.W., J.U. Farley, R.E. Lucas and L.W. Ring (1978): "Estimation and Inference for linear models in which subsets of the dependent variable are constrained," *Journal of the American Statistical Association*, 63, 1201-1213.

Newey, W.K. and K.D. West (1987): "A Simple Positive Semi-Definite, Heteroskedasiticity and Autocorrelation Consisitent Covariance Matrix," *Econometrica*, 55, 703-708.

Powell, A. (1969): "Aitken estimators as a tool in allocating predetermined aggregates," *Journal of the American Statistical Association*, 64, 913-922.

Theil, H. (1971): *Principles of Econometrics,* New York: John Wiley and Sons.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.

Table 1. Classical SUR System

Time	J_1	p value
Panel A: VWRETD with no adding up restrictions		
	Five year period	
1/65-12/69	23.47	0.009
1/70-12/74	10.46	0.401
1/75-12/79	25.78	0.004
1/80-12/84	20.20	0.027
1/85-12/89	15.65	0.110
1/90-12/94	8.07	0.621
	Ten year period	
1/65-12/74	15.01	0.132
1/75-12/84	22.63	0.012
1/84-12/94	15.15	0.127
	Thirty year period	
1/65-12/94	20.07	0.287
Panel B: Exact market portfolio with one equation deleted		
	Five year period	
1/65-12/69	19.78	0.019
1/70-12/74	9.81	0.366
1/75-12/79	25.67	0.002
1/80-12/84	9.04	0.434
1/85-12/89	12.46	0.189
1/90-12/94	5.85	0.755
	Ten year period	
1/65-12/74	14.98	0.092
1/75-12/84	16.72	0.053
1/85-12/94	8.42	0.492
	Thirty year period	
1/65-12/94	14.94	0.093

TABLE 2. SUR System with Heteroskedasticity and Autocorrelation

	J_3		J_4			
Time	Lag 0	p value	Lag 1	p value	Lag 4	p value
Panel A: VWRETD with no adding up restrictions						
Five year period						
1/65-12/69	24.02	0.007	33.78	0.000	45.24	0.000
1/70-12/74	10.31	0.410	10.66	0.384	19.67	0.032
1/75-12/79	33.03	0.000	41.89	0.000	49.08	0.000
1/80-12/84	21.02	0.021	22.48	0.013	25.02	0.005
1/85-12/89	17.89	0.057	22.46	0.013	37.59	0.000
1/90-12/94	8.63	0.567	8.99	0.532	19.72	0.032
Ten year period						
1/65-12/74	15.39	0.118	17.44	0.065	19.73	0.032
1/75-12/84	26.49	0.003	26.30	0.003	25.52	0.004
1/85-12/94	15.78	0.106	16.37	0.089	22.91	0.011
Thirty year period						
1/65-12/94	21.11	0.020	20.42	0.026	21.57	0.017
Panel B: Exact market portfolio with one equation deleted						
Five year period						
1/65-12/69	19.98	0.018	30.68	0.000	41.80	0.000
1/70-12/74	9.69	0.376	10.54	0.308	19.25	0.023
1/75-12/79	32.77	0.000	40.02	0.000	49.34	0.000
1/80-12/84	9.20	0.419	11.29	0.257	14.13	0.118
1/85-12/89	14.77	0.098	17.39	0.043	34.82	0.000
1/90-12/94	6.20	0.719	5.37	0.801	6.64	0.675
Ten year period						
1/65-12/74	15.37	0.081	17.43	0.042	19.49	0.021
1/75-12/84	16.84	0.051	17.12	0.047	18.25	0.032
1/85-12/94	9.42	0.399	9.60	0.383	11.80	0.225
Thirty year period						
1/65-12/94	15.47	0.079	15.81	0.071	17.20	0.046