Granger Centre
for time series econometrics

Granger Centre Discussion Paper Series

Size corrected significance tests in Seemingly Unrelated Regressions with autocorrelated errors

by

Spyridon D. Symeonides, Yiannis Karavias and Elias Tzavalis

Granger Centre Discussion Paper No. 14/01

The University of Nottingham

Size corrected significance tests in Seemingly Unrelated Regressions with autocorrelated errors

Spyridon D. Symeonides* Yiannis Karavias† Elias Tzavalis‡

November 5, 2013

Abstract

Refined asymptotic methods are used to produce degrees-of-freedom adjusted Edgeworth and Cornish-Fisher size corrections of the t and F testing procedures for the parameters of a S.U.R. model with serially correlated errors. The corrected tests follow the Student-t and F distributions, respectively, with an approximation error of order $O(\tau^3)$, where $\tau = 1/\sqrt{T}$ and T is the number of time observations. Monte Carlo simulations provide evidence that the size corrections suggested hereby have better finite sample properties, compared to the asymptotic testing procedures (either standard or Edgeworth corrected), which do not adjust for

*Department of Economics, University of Ioannina, Ioannina 455 00, Greece. *Email*: ssymeoni@cc.uoi.gr

†Corresponding Author, School of Economics, University of Nottingham and Granger Centre for Time Series Econometrics. *Email*: ioannis.karavias@nottingham.ac.uk

‡Department of Economics, Athens University of Economics and Business, 76 Patission Str., 10434 Athens, Greece. *Email*: etzavalis@aueb.gr

the degrees of freedom.

Key words: Linear regression; S.U.R. models; stochastic expansions; asymptotic approximations; AR(1) errors.

JEL classification: C10, C12, D24.

1 Introduction

Refined asymptotic methods can considerably improve the finite-sample performance of estimation and testing procedures in applied econometric research (see, e.g., Ullah (2004), for a survey). These methods involve higher order asymptotic approximations of the distributions of well known econormertic estimators and/or test statistics which can efficiently approximate their sample distributions (see Magdalinos and Symeonides (1995), Magee (1985), Rothenberg (1984b), Symeonides et al. (2007), *inter alia*). In finite samples, considerable discrepancies between the true and estimated values of these estimators or normality of error terms lead to substantial differences between the actual (sample) and nominal size of standard testing procedures. In the literature, these discrepancies are found to be very severe, especially for the linear regression model with non-scalar covariance matrix of error terms estimated by the feasible generalized least squares (FGLS), or maximum likelihood. Estimation of this model requires efficient methods of estimating the nuisance parameters of the error term covariance matrix.

Despite the substantial amount of work on refined asymptotic bias expansions of alternative estimators for the linear regression model or simultaneous systems of equations (see, e.g., Iglesias and Phillips (2010, 2011), Kiviet and Phillips (1996), Phillips (2000, 2007), *inter alia*), there are only a few papers in the literature of applying these methods to conventional test statistics, like the F and t ones. Rothenberg (1984b, 1988) used Edgeworth expansions in terms of the chi-square and normal distributions, respectively, to derive general formulae of corrected critical values of the Wald (or F) and t test statistics. Instead of using Edgeworth corrections of the critical values, Magdalinos and Symeonides (1995) suggested the use of degrees-of-freedom-adjusted Cornish-Fisher corrected t and F statistics for the linear regression model with first-order autocorrelated errors. Cornish-Fisher corrected t and F statistics for the linear regression model with heteroskedastic error term have been recently suggested by Symeonides et al (2007).

In this paper, we extend the above testing procedures of Rothemberg, Magdalinos and Symeonides to systems of Seemingly Unrelated Regression (S.U.R.) equations which allow for first-order autoregressive error terms. This is a multi-regression model which is frequently used in economics to simultaneously estimate investment functions, arbitrage asset pricing models, demand equations of different economic units (like industries, assets or countries) allowing for cross-correlation among them. Panel data models with fixed or random effects can be seen as special cases of the S.U.R model. Allowing for autoregression in the error terms, the S.U.R model can capture dynamic effects of the dependent and independent variables on economic relationships of interest, often considered in macroeconomic studies.

In particular, the paper derives degrees-of-freedom adjusted Edgeworth corrected critical values and Cornish-Fisher corrected statistics of the t and F testing procedures, for the above extension of the S.U.R. model, with serially correlated errors, estimated using FGLS. These corrections follow the Student-t

and F distributions, respectively, with an approximation error of order $O(\tau^3)$, where $\tau = 1/\sqrt{T}$ and T is the number of time observations of the sample. The use of degree-of-freendom adjusted forms of the above tests lead to approximations that are 'locally exact', i.e., the approximate distributions reduce to the exact ones, when the model is sufficiently simplified Magdalinos (1985). These approximations are found to increase the small sample performance of the tests (see Magdalinos and Symeonides (1995), Symeonides et al. (2007)). To our knowledge, this is the first paper in the literature which suggests size corrected test statistics for the S.U.R. model with serially correlated errors. The most closely related to our work is that of Strivastava and Maekawa (1995) who provided an Edgeworth expansion of the limiting distribution of the FGLS estimator of the S.U.R. model under the assumption of non-normal error terms.

Since the Cornish-Fisher expansion is the inversion of the Edgeworth correction of the critical values, the Edgeworth and Cornish-Fisher size corrections are asymptotically equivalent to the order of the required accuracy. However, the use of the Cornish-Fisher corrected test statistics, instead of the Edgeworth corrected critical values, can be recommended, in practice, for the following two main reasons (see Cornish and Fisher (1937), Fisher and Cornish (1960), Hill and Davis (1968), Magdalinos (1985), *inter alia*): First, they are proper random variables and their distributions have well-behaved tails, whereas the Edgeworth approximations are not well-defined distribution functions. The latter may assign negative 'probabilities' in the tails of the distributions. Second, the Cornish-Fisher corrected test statistics can be readily implemented in applied research based on the tables of standard distributions, which are publicly available. They do not require the calculation of new critical values.

The paper is organised as follows. Section 2 provides some preliminary notations. Section 3 presents the S.U.R. model and the assumptions needed in our expansions. Analytic formulae for the locally exact Edgeworth and Cornish-Fisher second order size corrections of the t and F test statistics are derived in Section 4. Section 5 conducts out a Monte Carlo exercise evaluating the performance of the suggested corrected tests. Finally, Section 6 concludes the paper. Proofs of the results of the paper are given in the Appendix.

2 Preliminary notation

Throughout the paper, we use the tr, vec, $\otimes$, and matrix differentiation notation as defined in Dhrymes (1978, pages 518–540), and for any two indices i and j, we denote Kronecker's delta as δ_{ij}. Moreover, any $(n \times m)$ matrix L with elements l_{ij} is denoted as

$$L = [(l_{ij})_{i=1,\,\ldots,\,n;\; j=1,\,\ldots,\,m}],$$

with obvious modifications for vectors and square matrices. If l_{ij} are $(n_i \times m_j)$ matrices, then L is the $(\sum n_i \times \sum m_j)$ partitioned matrix with submatrices l_{ij}. The following matrices:

$$P_X = X(X'X)^{-1}X', \;\; \overline{P}_X = I - P_X = I - X(X'X)^{-1}X'$$

denote the orthogonal projectors into the spaces spanned by the columns of the matrix X and its orthogonal complement, respectively. Finally, for any stochastic quantity (scalar, vector, or matrix) we use the symbol $\mathcal{E}(\cdot)$ to denote the expectation operator.

3 The model

Consider a S.U.R. system of M contemporaneously correlated regression equations of the form

$$y_\mu = X_\mu \beta_\mu + u_\mu \quad (\mu = 1, \ldots, M), \tag{1}$$

where y_μ are $(T \times 1)$ vectors of observations on the dependent variables, X_μ are $(T \times n_\mu)$ matrices of observations on sets of n_μ non-stochastic regressors, β_μ are $(n_\mu \times 1)$ vectors of parameters to be estimated and u_μ are $(T \times 1)$ vectors of non-observable serially correlated stochastic error terms of the μ-th equation, defined as $u_{t\mu}$ $(t = 1, \ldots, T)$. These terms are generated by the following stationary first-order autoregressive (AR(1)) process:

$$u_{t\mu} = \rho_\mu u_{(t-1)\mu} + \varepsilon_{t\mu}, \quad -1 < \rho_\mu < 1 \quad (t = 1, \ldots, T;\ \mu = 1, \ldots, M), \tag{2}$$

where $\varepsilon_{t\mu}$ are normally distributed innovations. For any two indices $\mu,\ \mu' = 1, \ldots, M$, we have $\mathcal{E}(\varepsilon_{t\mu}) = 0$, for all t. Moreover, for $t \neq 1$ or $t' \neq 1$, the covariance between two innovations $\varepsilon_{t\mu}$ and $\varepsilon_{t'\mu'}$ is given as $\mathcal{E}(\varepsilon_{t\mu}\varepsilon_{t'\mu'}) = \delta_{tt'}\sigma_{\mu\mu'}$. For $t = t' = 1$ and $\mu,\ \mu' = 1, \ldots, M$, $\mathcal{E}(\varepsilon_{t\mu}\varepsilon_{t'\mu'})$ becomes

$$\mathcal{E}(\varepsilon_{1\mu}\varepsilon_{1\mu'}) = \sigma_{\mu\mu'}(1-\rho_\mu^2)^{1/2}(1-\rho_{\mu'}^2)^{1/2}/(1-\rho_\mu\rho_{\mu'}), \tag{3}$$

see Parks (1967, pages 507–508). In addition to assumption $\rho_\mu \in (-1, 1)$, stationarity of AR(1) processes (2) implies the following relationships on the initial conditions of the error terms of the S.U.R. equations:

$$u_{1\mu} = (1-\rho_\mu^2)^{-1/2}\varepsilon_{1\mu} \quad (t = 1;\ \mu = 1, \ldots, M). \tag{4}$$

These relationships imply that, for all $t = 1, \ldots, T$ and $\mu, \mu' = 1, \ldots, M$, the error terms $u_{t\mu}$ satisfy the following conditions:

$$\mathcal{E}(u_{t\mu}) = 0, \quad \mathcal{E}(u_{t\mu}^2) = \sigma_{\mu\mu}/(1-\rho_\mu^2), \quad \mathcal{E}(u_{t\mu}u_{t\mu'}) = \sigma_{\mu\mu'}/(1-\rho_\mu\rho_{\mu'}). \tag{5}$$

Let $n = \sum_{\mu=1}^{M} n_\mu$, and define the $(MT \times 1)$ vectors y and u, the $(n \times 1)$ vector β and the $(MT \times n)$ block diagonal matrix X as follows:

$$\begin{gathered} y = [(y_\mu)_{\mu=1,\ldots,M}], \quad u = [(u_\mu)_{\mu=1,\ldots,M}], \\ \beta = [(\beta_\mu)_{\mu=1,\ldots,M}], \\ X = [(\delta_{\mu\mu'}X_\mu)_{\mu,\mu'=1,\ldots,M}]. \end{gathered} \tag{6}$$

Then, the system of equations (1) can be written in a matrix form as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{bmatrix}, \tag{7}$$

or, more compactly in a vectorized form, as

$$y = X\beta + u. \tag{8}$$

To derive size corrected significance tests for the elements of the vector β, the above representations of the S.U.R. system will be written in an autocorrelation-free form, after applying appropriate transformations on y, X and u. Following

Parks (1967), define the $(T \times T)$ matrices P_μ and $R^{\mu\mu'}$ as follows:

$$P_\mu = \begin{bmatrix} (1-\rho_\mu^2)^{-\frac{1}{2}} & 0 & 0 & \cdots & 0 \\ (1-\rho_\mu^2)^{-\frac{1}{2}}\rho_\mu & 1 & 0 & \cdots & 0 \\ (1-\rho_\mu^2)^{-\frac{1}{2}}\rho_\mu^2 & \rho_\mu & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (1-\rho_\mu^2)^{-\frac{1}{2}}\rho_\mu^{T-1} & \rho_\mu^{T-2} & \rho_\mu^{T-3} & \cdots & 1 \end{bmatrix}, \quad R^{\mu\mu'} = P_\mu^{-1\prime} P_{\mu'}^{-1}, \qquad (9)$$

and the following $(MT \times MT)$ block diagonal matrix

$$P = [(\delta_{\mu\mu'} P_\mu)_{\mu,\mu'=1,\,\dots,\,M}]. \qquad (10)$$

Then, (2) implies that the $(T \times 1)$ random vectors u_μ can be written as

$$u_\mu = P_\mu \varepsilon_\mu \quad (\mu = 1,\, \dots,\, M), \qquad (11)$$

where ε_μ are $(T \times 1)$ random vectors with non-autocorrelated elements $\varepsilon_{t\mu}$, i.e.,

$$\varepsilon_\mu = [(\varepsilon_{t\mu})_{t=1,\,\dots,\,T;\;\mu=1,\,\dots,\,M}]. \qquad (12)$$

As in (11), consider the $(T \times 1)$ vectors $y_{\mu *}$ and $(T \times n_\mu)$ matrices $X_{\mu *}$, with non-autocorrelated elements, satisfying the following relations:

$$y_{\mu *} = P_\mu^{-1} y_\mu, \quad X_{\mu *} = P_\mu^{-1} X_\mu, \qquad (13)$$

and define the $(MT \times 1)$ vector y_* and $(MT \times n)$ block diagonal matrix X_* as follows:

$$y_* = [(y_{\mu *})_{\mu=1,\,\dots,\,M}], \quad X_* = [(\delta_{\mu\mu'} X_{\mu *})_{\mu,\mu'=1,\,\dots,\,M}]. \qquad (14)$$

Then, premultiplying the μ-th equation of (7) by P_μ^{-1}, we can derive the following S.U.R. model with non-autocorrelated error terms:

$$\begin{bmatrix} y_{1*} \\ y_{2*} \\ \vdots \\ y_{M*} \end{bmatrix} = \begin{bmatrix} X_{1*} & 0 & \cdots & 0 \\ 0 & X_{2*} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_{M*} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix} \tag{15}$$

(see Zellner (1962, 1963), Zellner and Huang (1962), Zellner and Theil (1962)). In more compact form, this model can be written as

$$y_* = X_*\beta + \varepsilon, \tag{16}$$

where $y_* = P^{-1}y$, $X_* = P^{-1}X$ and $\varepsilon = P^{-1}u$. The above representation of the S.U.R. system implies that the $(MT \times 1)$ error vector u in (8) is normally distributed with mean and variance-covariance matrix given as follows:

$$\mathcal{E}(u) = 0, \quad \mathcal{E}(uu') = \Omega^{-1} = P\mathcal{E}(\varepsilon\varepsilon')P' = P(\Sigma \otimes I_T)P', \tag{17}$$

where

$$\Sigma = [(\sigma_{\mu\mu'})_{\mu,\mu'=1,\,\ldots,\,M}]. \tag{18}$$

The last relationship implies that

$$\Omega = P'^{-1}(\Sigma^{-1} \otimes I_T)P^{-1} \tag{19}$$

is a function of the $((M + M^2) \times 1)$ parameter vector $\gamma = (\varrho', \varsigma')'$, where $\varrho = (\rho_1, \ldots, \rho_M)'$ is the $(M \times 1)$ vector of autocorrelation coefficients in (2) and ς is the $(M^2 \times 1)$ vector $\varsigma = vec(\Sigma^{-1}) \in \pounds = \mathbb{R}^{M^2} - \mho$, where $\mho$ is the subspace of $\mathbb{R}^{M^2}$ in which Σ is not positive definite. After defining the composite index

$$(\mu\mu') = \mu + M(\mu' - 1) \quad ((\mu\mu') = 1, \ldots, M^2), \tag{20}$$

for any two indices $\mu, \mu' = 1, \ldots, M$, it can be easily seen that the $(\mu\mu')$-th element of vector ς, denoted as $\varsigma_{(\mu\mu')}$, is actually the (μ, μ')-th element of matrix Σ^{-1}, denoted as $\sigma^{\mu\mu'}$.

The system of equations (16) (or (15)) can be seen as the vectorized representation of the following form of the S.U.R. model of M equations:

$$Y_* = ZB + E, \tag{21}$$

where Y_* and E are $(T \times M)$ random matrices defined as

$$y_* = vec(Y_*), \quad \varepsilon = vec(E), \tag{22}$$

respectively, where the rows of matrix E are $\mathcal{N}_M(0, \Sigma)$ random vectors and B is a $(K \times M)$ matrix whose columns, denoted as b_μ, are defined as

$$b_\mu = \Psi_\mu \beta_\mu \quad (\mu = 1, \ldots, M), \tag{23}$$

where Ψ_μ are $(K \times n_\mu)$ known submatrices of the $(MK \times n)$ block diagonal matrix

$$\Psi = [(\delta_{\mu\mu'} \Psi_\mu)_{\mu,\mu'=1,\ldots,M}]. \tag{24}$$

Finally, Z is a $(T \times K)$ matrix with non-autocorrelated columns, defined by the following relationship:

$$\begin{aligned} X_* &= [(\delta_{\mu\mu'} X_{\mu*})_{\mu,\mu'=1,\ldots,M}] = [(\delta_{\mu\mu'} Z \Psi_\mu)_{\mu,\mu'=1,\ldots,M}] \\ &= [(\delta_{\mu\mu'} Z)_{\mu,\mu'=1,\ldots,M}][(\delta_{\mu\mu'} \Psi_\mu)_{\mu,\mu'=1,\ldots,M}] \\ &= (I_M \otimes Z)\Psi. \end{aligned} \tag{25}$$

The above representation of the S.U.R. model, given by (21), will facilitate the expansions needed in our derivations of the size corrected tests suggested in the paper.

3.1 Assumptions

To carry out our expansions, it would be theoretically convenient to introduce a reparameterization of the error covariance matrix of model (8) as follows:

$$y = X\beta + \sigma u, \quad \sigma > 0, \quad u \sim \mathcal{N}_{MT}(0, \Omega^{-1}), \tag{26}$$

assuming that parameter σ^2 can be estimated separately from the rest terms of the covariance matrix Ω^{-1} of vector u.[1]

For the derivation of our size correceted tests, we need to make a number of assumptions on the elements of matrix Ω, which is the inverse of the variance-covariance matrix of the error vector u. To this end, we denote as Ω_i, Ω_{ij}, etc., the $(MT \times MT)$ matrices of first, second and higher-order derivatives, respectively, of the elements of matrix Ω with respect to the elements of the $((M + M^2) \times 1)$ vector of nuisance parameters $\gamma = (\varrho', \varsigma')'$. For any estimator of γ, define the $((1 + M + M^2) \times 1)$ vector δ, with elements

$$\delta_0 = \frac{\hat{\sigma}^2 - 1}{\tau}, \quad \delta_{\rho_\mu} = \frac{\hat{\rho}_\mu - \rho_\mu}{\tau}, \quad \delta_{\varsigma_{(\mu\mu')}} = \frac{\hat{\varsigma}_{(\mu\mu')} - \varsigma_{(\mu\mu')}}{\tau}, \tag{27}$$

where $\mu = 1, \ldots, M$, $(\mu\mu') = 1, \ldots, M^2$ and $\tau = 1/\sqrt{T}$ is the 'asymptotic scale' of our second order stochastic expansions. Then, our size corrected tests can be derived based on the following assumption.

Assumption 1:

[1]The nuisance parameters σ and γ can be simultaneously identified under the restriction $\sigma = 1$, which implies that the estimate of matrix Σ, denoted as $\hat{\Sigma}$, is accurate, up to a multiplicative factor. This is not true in samples with small time dimension. A convenient method to estimate σ is through the following feasible GL estimator

$$\hat{\sigma}_{GL} = \left[(y - X\hat{\beta})' \left(\hat{P}'^{-1}_{GL}(\hat{\Sigma}^{-1}_{GL} \otimes I_T)\hat{P}^{-1}_{GL}\right)(y - X\hat{\beta})/(MT - n)\right]^{1/2},$$

where $\hat{\beta}$ is the feasible GL estimator based on any consistent estimators of Σ^{-1} and P^{-1}.

(i) The elements of Ω and Ω^{-1} are bounded for all T, all vectors ϱ with elements $\rho_\mu \in (-1, 1)$, and all vectors $\varsigma \in \pounds$. Moreover, the following matrices:

$$A = X'\Omega X/T, \quad F = X'X/T, \quad \Gamma = Z'Z/T \tag{28}$$

converge to non-singular limits, as $T \to \infty$.

(ii) Up to the fourth order, the partial derivatives of the elements of Ω with respect to the elements of ϱ and ς, are bounded for all T, all vectors ϱ with elements in the interval $(-1, 1)$, and all vectors $\varsigma \in \pounds$.

(iii) The estimators $\hat{\varrho}$ and $\hat{\varsigma}$ are even functions of u, and they are functionally unrelated to the parameter vector β, i.e., they can be written as functions of X, Z, and u only.

(iv) The vector of nuisance parameters δ admits a stochastic expansion of the form

$$\begin{aligned} \delta &= \left[\delta_0, \ [(\delta_{\rho_\mu})_{\mu=1,\ldots,M}]', \ [(\delta_{\varsigma_{(\mu\mu')}})_{(\mu\mu')=1,\ldots,M^2}]' \ \right]' \\ &= d_1 + \tau d_2 + \omega(\tau^2), \end{aligned} \tag{29}$$

where the order of magnitude $\omega(\cdot)$, defined in the Appendix, has the same operational properties as order $O(\cdot)$, and the expectations

$$\mathcal{E}(d_1 d_1'), \quad \mathcal{E}(\sqrt{T} d_1 + d_2) \tag{30}$$

exist and have finite limits, as $T \to \infty$.

The first two conditions of Assumption 1 imply that the following matrices:

$$A_i = X'\Omega_i X/T, \quad A_{ij} = X'\Omega_{ij} X/T, \quad A^*_{ij} = X'\Omega_i \Omega^{-1} \Omega_j X/T \tag{31}$$

are bounded. Thus, according to Magdalinos (1992), the Taylor series expansion of β constitutes a stochastic expansion. Since the vectors of nuisance parameters ϱ and ς are functionally unrelated to β, condition (iii) of Assumption 1 is satisfied for a wide class of estimators $\hat{\varrho}$ and $\hat{\varsigma}$, including the maximum likelihood estimators and the simple or iterative estimators based on the regression residuals (see Breusch (1980), Rothenberg (1984a)). Note that we *need not* assume that estimators $\hat{\varrho}$ and $\hat{\varsigma}$ are asymptotically efficient.

Moreover, conditions (i)–(iv) of Assumption 1 should be satisfied by all the estimators of ϱ and ς, considered in the paper. The estimators of the elements of ϱ, i.e., ρ_μ $(\mu = 1, ..., M)$ include the following: the least squares (LS), Durbin-Watson (DW), generalized least squares (GL), Prais-Winsten (PW) and maximum likelihood (ML).[2] The elements of vector $\varsigma = vec(\Sigma^{-1})$ can be estimated by

$$\hat{\varsigma} = vec\left[(Y_* - Z\hat{B})'(Y_* - Z\hat{B})/T\right]^{-1}, \tag{32}$$

where $\hat{B}$ is any consistent estimator of the matrix of parameters B of regression model (21). Consistent estimators of B include the unrestricted and restricted least squares (denoted as UL and RL, respectively), the simple and iterative generalized least squares (denoted as GL and IG, respectively) and the maximum likelihood (ML) estimators.[3]

To present the expansions suggested in the paper, expectations $\mathcal{E}(d_1 d_1')$ and

---

[2]The closed forms of these estimators of ρ_μ, for all μ, are given as follows:

(i) LS:

$$\tilde{\rho}_\mu = \sum_{t=2}^{T} \tilde{u}_{t\mu}\tilde{u}_{(t-1)\mu} \Big/ \sum_{t=1}^{T} \tilde{u}_{t\mu}^2,$$

where $\tilde{u}_{t\mu}$ are the LS residuals of regression model (1).

(ii) DW:

$$\hat{\rho}_\mu^{(DW)} = 1 - (DW/2),$$

where the DW is the Durbin-Watson statistic.

(iii) GL:

$$\hat{\rho}_\mu = \sum_{t=2}^{T} \hat{u}_{t\mu}\hat{u}_{(t-1)\mu} \Big/ \sum_{t=1}^{T} \hat{u}_{t\mu}^2,$$

where $\hat{u}_{t\mu}$ denote the GL estimates of $u_{t\mu}$, based on the autocorrelation-correction of regression model (1), for all μ, using any asymptotically efficient estimator of ρ_μ.

(iv) PW: This estimator of ρ_μ, denoted as $\hat{\rho}_\mu^{(PW)}$, together with the PW estimator of β, denoted as $\hat{\beta}_\mu^{(PW)}$, minimize the sum of squared GL residuals (Prais and Winsten (1954)).

(v) ML: This estimator, denoted as $\hat{\rho}_\mu^{(ML)}$, satisfies a cubic equation with coefficients defined in terms of the ML residuals (Beach and MacKinnon (1978)).

[3]The closed forms of these estimators of B are given as follows:

(i) UL:

$$\hat{B}_{(UL)} = (Z'Z)^{-1}Z'Y_*.$$

$\mathcal{E}(\sqrt{T}d_1 + d_2)$ will be defined as follows:

$$\lim_{T\to\infty} \mathcal{E}(d_1 d_1') = \begin{bmatrix} \lambda_0 & \lambda_\varrho' & \lambda_\varsigma' \\ \lambda_\varrho & \Lambda_\varrho & \Lambda_{\varrho\varsigma}' \\ \lambda_\varsigma & \Lambda_{\varrho\varsigma} & \Lambda_\varsigma \end{bmatrix} \quad \text{and} \quad \lim_{T\to\infty} \mathcal{E}(\sqrt{T}d_1 + d_2) = \begin{bmatrix} \kappa_0 \\ \kappa_\varrho \\ \kappa_\varsigma \end{bmatrix}, \tag{33}$$

respectively, where λ_0 and κ_0 are scalars, λ_ϱ and κ_ϱ are $(M \times 1)$ vectors, λ_ς and κ_ς are $(M^2 \times 1)$ vectors, Λ_ϱ is a $(M \times M)$ matrix, Λ_ς is a $(M^2 \times M^2)$ matrix and $\Lambda_{\varrho\varsigma}$ is a $(M^2 \times M)$ matrix. The following partitions of the above matrix and vector will be of use in the paper:

$$\begin{bmatrix} \lambda_0 & \lambda' \\ \lambda & \Lambda \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \kappa_0 \\ \kappa \end{bmatrix}, \tag{34}$$

where

$$\Lambda = \begin{bmatrix} \Lambda_\varrho & \Lambda_{\varrho\varsigma}' \\ \Lambda_{\varrho\varsigma} & \Lambda_\varsigma \end{bmatrix}, \quad \lambda = \begin{bmatrix} \lambda_\varrho \\ \lambda_\varsigma \end{bmatrix} \quad \text{and} \quad \kappa = \begin{bmatrix} \kappa_\varrho \\ \kappa_\varsigma \end{bmatrix}, \tag{35}$$

where Λ is a $((M + M^2) \times (M + M^2))$ matrix, and λ and κ are $((M + M^2) \times 1)$ vectors. The elements of the vectors and matrices in (33), (34) and (35) can be interpreted as 'measures' of the accuracy of the expansions of estimators $\hat{\sigma}^2$, $\hat{\rho}_\mu$ and $\hat{\varsigma}_{(\mu\mu')}$ around the true values of the corresponding parameters.

(ii) RL:

$$vec(\hat{B}_{(RL)}) = \Psi(X_*' X_*)^{-1} X_*' y_*.$$

(iii) GL:

$$vec(\hat{B}_{(GL)}) = \Psi \left[X_*'(\hat{\Sigma}_I^{-1} \otimes I_T) X_*\right]^{-1} X_*'(\hat{\Sigma}_I^{-1} \otimes I_T) y_*,$$

where $\hat{\Sigma}_I^{-1}$ is the UL or RL estimator of Σ.

(iv) IG: This estimator, denoted as $\hat{B}_{(IG)}$, is computed by iterative implementation of the GL estimator.

(v) ML: This estimator, denoted as $\hat{B}_{(ML)}$, can be computed by iterating the GL estimation process up to convergence (Dhrymes (1971)).

4 Size corrected test statistics

In this section, we derive size corrected t, Wald and F test statistics, as well as the second-order approximations of their distributions based on the conditions of Assumption 1. The versions of the test statistics which adjust for the degrees of freedom, namely the Student-t and F, are locally exact. That is, if the vector of parameters $\gamma = (\varrho', \varsigma')'$ is known to belong to a ball of radius ϑ, then the approximate distributions of these test statistics become exact, as $\vartheta \to 0$.

4.1 The t test

Let e be a $(n \times 1)$ vector of known quantities and e_0 be a known scalar. To test null hypothesis

$$H_0 : \; e'\beta = e_0 \tag{36}$$

against its one-sided alternatives, the t statistic takes the following form:

$$t = (e'\beta - e_0) / \left[\hat{\sigma}^2 e'(X'\hat{\Omega}X)^{-1} e\right]^{1/2}. \tag{37}$$

This statistic takes into account the degrees of freedom of the Student-t distribution.

For the derivation of the suggested asymptotic expansions, we define the $((M+M^2)\times 1)$ vector l and the $((M+M^2)\times(M+M^2))$ matrix L as follows:

$$l = \left[\; [(l_{\rho_\mu})_{\mu=1,\,\dots,\,M}]', \quad [(l_{\varsigma_{(\mu\mu')}})_{(\mu\mu')=1,\,\dots,\,M^2}]' \; \right]', \tag{38}$$

$$L = \begin{bmatrix} [(l_{\rho_\mu \rho_{\mu'}})_{\mu,\mu'=1,\dots,M;}] & [(l_{\rho_\mu \varsigma_{(\nu\nu')}})_{\substack{\mu=1,\,\dots,\,M; \\ (\nu\nu')=1,\dots,M^2}}] \\ & \\ [(l_{\varsigma_{(\nu\nu')}\rho_\mu})_{\substack{(\nu\nu')=1,\,\dots,\,M^2; \\ \mu=1,\,\dots,\,M}}] & [(l_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}})_{\substack{(\mu\mu')=1,\,\dots,\,M^2; \\ (\nu\nu')=1,\,\dots,\,M^2}}] \end{bmatrix}, \tag{39}$$

where the elements of the vector l and the matrix L are defined below:

$$
\begin{aligned}
l_{\rho_\mu} &= h'GA_{\rho_\mu}Gh, \quad l_{\varsigma_{(\mu\mu')}} = h'GA_{\varsigma_{(\mu\mu')}}Gh, \\
l_{\rho_\mu\rho_{\mu'}} &= h'GC_{\rho_\mu\rho_{\mu'}}Gh, \quad l_{\rho_\mu\varsigma_{(\nu\nu')}} = h'GC_{\rho_\mu\varsigma_{(\nu\nu')}}Gh, \qquad (40) \\
l_{\varsigma_{(\nu\nu')}\rho_\mu} &= h'GC_{\varsigma_{(\nu\nu')}\rho_\mu}Gh, \quad l_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}} = h'GC_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}}Gh,
\end{aligned}
$$

where $G = A^{-1} = (X'\Omega X/T)^{-1}$ is a $(n \times n)$ matrix, $h = e/(e'Ge)^{1/2}$ is a $(n \times 1)$ vector and

$$
\begin{aligned}
C_{\rho_\mu\rho_{\mu'}} &= A^*_{\rho_\mu\rho_{\mu'}} - 2A_{\rho_\mu}GA_{\rho_{\mu'}} + A_{\rho_\mu\rho_{\mu'}}/2, \\
C_{\rho_\mu\varsigma_{(\nu\nu')}} &= A^*_{\rho_\mu\varsigma_{(\nu\nu')}} - 2A_{\rho_\mu}GA_{\varsigma_{(\nu\nu')}} + A_{\rho_\mu\varsigma_{(\nu\nu')}}/2, \qquad (41) \\
C_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}} &= A^*_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}} - 2A_{\varsigma_{(\mu\mu')}}GA_{\varsigma_{(\nu\nu')}} + A_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}}/2,
\end{aligned}
$$

with obvious modifications for $C_{\varsigma_{(\nu\nu')}\rho_\mu}$.

The next two theorems give Edgeworth approximations of the distribution functions of the t statistic, given by (37), and its version which adjusts for the degrees of freedom.

Theorem 1. *Under null hypothesis (36), the distribution function of the t statistic, given by (37), admits the Edgeworth expansion*

$$
\Pr\{t \le x\} = I(x) - \frac{\tau^2}{2}\left[\left(p_1 + \tfrac{1}{2}\right) + \left(p_2 + \tfrac{1}{2}\right)x^2\right] x i(x) + O(\tau^3), \qquad (42)
$$

where $I(\cdot)$ and $i(\cdot)$ are the standard normal distribution and density functions, respectively, and

$$
p_1 = tr(\Lambda L) + \frac{l'\Lambda l}{4} + l'(\kappa + \frac{\lambda}{2}) - \kappa_0 + \frac{\lambda_0 - 2}{4}, \quad p_2 = \frac{l'\Lambda l - 2l'\lambda + \lambda_0 - 2}{4}. \qquad (43)
$$

Analytic formulae for the computation of scalars λ_0, κ_0, and the elements of λ, κ, Λ, l and L are given in the Appendix (see Lemmas A.15 and A.17).

Theorem 2. *Under null hypothesis (36), the distribution function of the t statistic, given by (37), admits the Edgeworth expansion*

$$\Pr\{t \leq x\} = I_{MT-n}(x) - \frac{\tau^2}{2}\left[p_1 + p_2 x^2\right] x i_{MT-n}(x) + O(\tau^3), \tag{44}$$

where $I_{MT-n}(\cdot)$ and $i_{MT-n}(\cdot)$ are the Student-t distribution and density functions, respectively, and quantities p_1 and p_2 are defined in (43).

Theorem 1 implies that we can calculate the Edgeworth corrected $\alpha\%$ critical value of t statistic (37) as

$$n_\alpha^* = n_\alpha + \frac{\tau^2}{2}\left[\left(p_1 + \tfrac{1}{2}\right) + \left(p_2 + \tfrac{1}{2}\right) n_\alpha^2\right] n_\alpha, \tag{45}$$

based on the $\alpha\%$ significant point of the standard normal distribution, denoted as n_α. Similarly, based on Theorem 2, we can calclulate the Edgeworth corrected $\alpha\%$ critical value of t statistic (37) as

$$t_\alpha^* = t_\alpha + \frac{\tau^2}{2}\left[p_1 + p_2 t_\alpha^2\right] t_\alpha, \tag{46}$$

using the $\alpha\%$ significant point of the Student-t distribution, denoted as t_α.

The Edgeworth approximation employed by Theorems 1 and 2 to obtain the size corrected critical values n_α^* and t_α^* is not a proper distribution function, as it can assign negative 'probabilities' in the tails of the approximate distribution. To overcome this problem, we can use a Cornish-Fisher expansion. This corrects the test statistics of interest, instead of their critical values. The Cornish-Fisher expansion is simply the inversion of the Edgeworth correction of the critical values and, thus, it is expected to have very similar properties around the mean of the approximate distribution. However, at the tails of this distribution, which are important for inference, the properties of the Cornish-Fisher expansion are different. In fact, the Cornish-Fisher size corrected statistics constitute random variables with well-behaved tails, and thus they *do not* assign negative 'probabilities' at the tails of their distributions.

The Cornish-Fisher corrected t statistic for testing null hypothesis (36) is given in the following theorem.

Theorem 3. *Under null hypothesis (36), the Cornish-Fisher size corrected t statistic*

$$t_* = t - \frac{\tau^2}{2}\left[p_1 + p_2 t^2\right] t \tag{47}$$

is distributed, with an approximation error of order $O(\tau^3)$, as a Student-t random variable with $MT - n$ degrees of freedom.

The Cornish-Fisher size corrected t statistic t_*, given by equation (47), can be readily used, in practice, to test null hypothesis (36) against its one-sided alternatives. This can be done by using the tables of the Student-t distribution with $MT - n$ degrees of freedom.

4.2 The Wald and F tests

Let H be a known $(m \times n)$ matrix of rank m and h_0 be a known $(m \times 1)$ vector. To test null hypothesis

$$H_0 : \; H\beta = h_0, \tag{48}$$

against all possible alternatives, we can use the Wald statistic

$$w = (H\hat{\beta} - h_0)' \left[H(X'\hat{\Omega}X)^{-1}H'\right]^{-1} (H\hat{\beta} - h_0)/\hat{\sigma}^2, \tag{49}$$

or the familiar F statistic

$$F = (H\hat{\beta} - h_0)' \left[H(X'\hat{\Omega}X)^{-1}H'\right]^{-1} (H\hat{\beta} - h_0)/m\hat{\sigma}^2, \tag{50}$$

which adjusts for the degrees of freedom.

For the derivation of the suggested asymptotic expansions, we define the $(n \times n)$ matrix

$$Q = H'(HGH')^{-1}H, \tag{51}$$

and we partition the $(n \times n)$ matrices $G = A^{-1} = (X'\Omega X/T)^{-1}$ and $\Xi = GQG$ and the $(n \times 1)$ vector h as follows:

$$G = [(G_{ij})_{i,j=1,\,\ldots,\,M}], \quad \Xi = [(\Xi_{ij})_{i,j=1,\,\ldots,\,M}], \quad h = [(h_i)_{i=1,\,\ldots,\,M}], \tag{52}$$

where G_{ij} and Ξ_{ij} are the (i,j)-th $(n_i \times n_j)$ submatrices of G and Ξ, respectively, and $h_i = e_i/(e'Ge)^{1/2}$ is the i-th $(n_i \times 1)$ subvector of h, where e_i is the corresponding i-th $(n_i \times 1)$ subvector of the $(n \times 1)$ vector e.

Next, define the $((M+M^2)\times 1)$ vector c, and the $((M+M^2)\times(M+M^2))$ matrices C and D_* as follows:

$$c = \Big[\; [(c_{\rho_\mu})_{\mu=1,\,\ldots,\,M}]', \quad [(c_{\varsigma_{(\mu\mu')}})_{(\mu\mu')=1,\,\ldots,\,M^2}]' \;\Big]', \tag{53}$$

$$C = \begin{bmatrix} [(c_{\rho_\mu\rho_{\mu'}})_{\mu,\mu'=1,\,\ldots,\,M}] & [(c_{\rho_\mu\varsigma_{(\nu\nu')}})_{\substack{\mu=1,\,\ldots,\,M;\\(\nu\nu')=1,\,\ldots,\,M^2}}] \\ [(c_{\varsigma_{(\nu\nu')}\rho_\mu})_{\substack{(\nu\nu')=1,\,\ldots,\,M^2;\\ \mu=1,\,\ldots,\,M}}] & [(c_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}})_{\substack{(\mu\mu')=1,\,\ldots,\,M^2;\\(\nu\nu')=1,\,\ldots,\,M^2}}] \end{bmatrix} \tag{54}$$

and

$$D_* = \begin{bmatrix} [(d_{\rho_\mu\rho_{\mu'}})_{\mu,\mu'=1,\,\ldots,\,M}] & [(d_{\rho_\mu\varsigma_{(\nu\nu')}})_{\substack{\mu=1,\,\ldots,\,M;\\(\nu\nu')=1,\,\ldots,\,M^2}}] \\ [(d_{\varsigma_{(\nu\nu')}\rho_\mu})_{\substack{(\nu\nu')=1,\,\ldots,\,M^2;\\ \mu=1,\,\ldots,\,M}}] & [(d_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}})_{\substack{(\mu\mu')=1,\,\ldots,\,M^2;\\(\nu\nu')=1,\,\ldots,\,M^2}}] \end{bmatrix}, \tag{55}$$

where the elements of vector c, and matrices C and D_* are defined as follows:

$$
\begin{aligned}
c_{\rho_\mu} = tr(A_{\rho_\mu}\Xi), \quad c_{\rho_\mu\rho_{\mu'}} = tr(C_{\rho_\mu\rho_{\mu'}}\Xi),\\
c_{\rho_\mu\varsigma_{(\nu\nu')}} = tr(C_{\rho_\mu\varsigma_{(\nu\nu')}}\Xi),\\
c_{\varsigma_{(\mu\mu')}} = tr(A_{\varsigma_{(\mu\mu')}}\Xi), \quad c_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}} = tr(C_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}}\Xi),\\
d_{\rho_\mu\rho_{\mu'}} = tr(D_{*\rho_\mu\rho_{\mu'}}\Xi), \quad d_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}} = tr(D_{*\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}}\Xi),\\
d_{\rho_\mu\varsigma_{(\nu\nu')}} = tr(D_{*\rho_\mu\varsigma_{(\nu\nu')}}\Xi),
\end{aligned}
\tag{56}
$$

where

$$
\begin{aligned}
D_{*\rho_\mu\rho_{\mu'}} = \frac{A_{\rho_\mu}\Xi A_{\rho_{\mu'}}}{2}, \quad D_{*\rho_\mu\varsigma_{(\nu\nu')}} = \frac{A_{\rho_\mu}\Xi A_{\varsigma_{(\nu\nu')}}}{2},\\
D_{*\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}} = \frac{A_{\varsigma_{(\mu\mu')}}\Xi A_{\varsigma_{(\nu\nu')}}}{2},
\end{aligned}
\tag{57}
$$

with obvious modifications for $c_{\varsigma_{(\nu\nu')}\rho_\mu}$, $d_{\varsigma_{(\nu\nu')}\rho_\mu}$ and $D_{*\varsigma_{(\nu\nu')}\rho_\mu}$.

The next two theorems give Edgeworth approximations of the distribution functions of the Wald (w) and F statistics, given by (49) and (50), respectively.

Theorem 4. *Under null hypothesis (48), the distribution function of Wald statistic w, given by (49), admits the Edgeworth expansion*

$$
\Pr\{w \le x\} = F_m(x) - \tau^2\,[\xi_1 + (\xi_2/(m+2))x]\,\frac{x}{m} f_m(x) + O(\tau^3), \tag{58}
$$

where $F_m(\cdot)$ and $f_m(\cdot)$ are the chi-square distribution and density functions, respectively, and

$$
\begin{aligned}
\xi_1 = tr[\Lambda(C + D_*)] - c'\Lambda c/4 + c'\kappa + m[c'\lambda/2 - \kappa_0 - (m-2)\lambda_0/4],\\
\xi_2 = tr(\Lambda D_*) + [c'\Lambda c - (m+2)(2c'\lambda - m\lambda_0)]/4.
\end{aligned}
\tag{59}
$$

Analytic formulae for the computation of scalars λ_0 and κ_0, and the elements of λ, κ, Λ, c, C and D_* are given in the Appendix (see Lemmas A.16 and A.17).

Theorem 5. *Under null hypothesis (48), the distribution function of F statistic, given by (50), admits the Edgeworth expansion*

$$\Pr\{F \leq x\} = F^m_{MT-n}(x) - \tau^2 \left[q_1 + q_2 x\right] x f^m_{MT-n}(x) + O(\tau^3), \tag{60}$$

where $F^m_{MT-n}(\cdot)$ and $f^m_{MT-n}(\cdot)$ are the F distribution and density functions, respectively, and

$$q_1 = \xi_1/m + (m-2)/2, \quad q_2 = \xi_2/(m+2) - m/2, \tag{61}$$

where quantities ξ_1 and ξ_2 are defined in (59).

Theorem 4 implies that the Edgeworth corrected $\alpha\%$ critical value of the Wald statistic (49) is given as

$$\chi^*_\alpha = \chi_\alpha + \tau^2 \left[\frac{\xi_1}{m} + \frac{\xi_2}{m(m+2)}\chi_\alpha\right]\chi_\alpha, \tag{62}$$

based on the $\alpha\%$ significant point of the chi-square distribution, denoted as χ_α. Theorem 5 enables us to calclulate the Edgeworth corrected $\alpha\%$ critical value of F statistic (50) as

$$F^*_\alpha = F_\alpha + \tau^2 \left[q_1 + q_2 F_\alpha\right] F_\alpha, \tag{63}$$

based on the $\alpha\%$ significant point of the F distribution, denoted as F_α.

The Cornish-Fisher size corrected F statistic for testing null hypothesis (48) is given in the next theorem.

Theorem 6. *Under null hypothesis (48), the Cornish-Fisher size corrected F statistic*

$$F_* = F - \tau^2 \left[q_1 + q_2 F\right] F \tag{64}$$

is distributed, with an approximation error of order $O(\tau^3)$, as an F random variable with m and $MT - n$ degrees of freedom.

Unlike the Edgeworth approximation, the Cornish-Fisher corrected F statistic, denoted as F_* in equation (64), is a proper random variable and it does not assign negative 'probabilities' in the tails of its distribution. Thus, the Cornish-Fisher corrected F statistic can be be readily implemented, in applied research, to test null hypothesis (48). This can be done by using the tables of the F distribution, with $MT - n$ degrees of freedom.

5 Monte-Carlo simulations

In this section, we evaluate the small-sample performance of the size corrected tests suggested in the previous section, compared to their corresponding standard (first-order asymptotic approximation) versions. To this end, we rely on a Monte Carlo simulation exercise based on 5000 iterations and we consider small-smaples of $T = 15, 20, 40$ observations.

In our simulation exercise, we consider the original S.U.R. model of $M = 2$ unrelated equations (see, e.g., Zellner (1962)), i.e.,

$$\begin{aligned} y_{t,1} &= \beta_{0,1} + \beta_{1,1}x_{t1,1} + \beta_{2,1}x_{t2,1} + u_{t,1} \\ y_{t,2} &= \beta_{0,2} + \beta_{1,2}x_{t1,2} + \beta_{2,2}x_{t2,2} + u_{t,2} \end{aligned} \qquad (t = 1, \ldots, T), \tag{65}$$

where error terms $u_{t,1}$ and $u_{t,2}$ are contemporaneously correlated with covariance σ_{12}. Both of these error terms follow AR(1) process (2), with normally distributed innovations. The autoregressive coefficients of this process ρ_1 and ρ_2 are assumed to be equal, i.e., $\rho_1 = \rho_2 = \rho \in (-1, 1)$. To ensure stationarity of error terms $u_{t,1}$ and $u_{t,2}$, conditions (3) are satisfied. For $t = 0$, these conditions require

$$y_{0,1} \sim \mathcal{N}(0, \sigma_{11}/(1-\rho_1^2)) \qquad \text{and} \quad \mathcal{E}(y_{0,1}y_{0,2}) = \sigma_{12}\frac{(1-\rho_1^2)^{1/2}(1-\rho_2^2)^{1/2}}{1-\rho_1\rho_2}.$$
$$y_{0,2} \sim \mathcal{N}(0, \sigma_{22}/(1-\rho_2^2))$$

In our analysis, we assume $\sigma_{11} = \sigma_{22} = 1$ and we are focused on investigating the consequences of the different sign and magnitude of covariances σ_{12} on our tests, for the following cases: $\sigma_{12} = \pm 0.1$, ± 0.5, ± 0.75, ± 0.9. Since $\sigma_{11} = \sigma_{22} = 1$, σ_{12} is the correlation coefficient between $u_{t,1}$ and $u_{t,2}$.

According to (15) (or (16)), the above S.U.R. model can be written in terms of the following transformed equations, with non-autocorrelated errors:

$$y_{1*} = X_{1*}\beta_1 + \varepsilon_1; \quad y_{2*} = X_{2*}\beta_2 + \varepsilon_2,$$

where y_{1*} and y_{2*} are $(TX1)$ vectors of observations on the dependent variables, with $P_\mu y_{\mu*} = y_\mu$, for $\mu = 1, 2$, where P_μ is defined by (9), X_{1*} and X_{2*} are $(T \times 3)$ matrices of regressors, with $P_\mu X_{\mu*} = X_\mu$ and $\beta_1 = (\beta_{0,1},\ \beta_{1,1},\ \beta_{2,1})'$, $\beta_2 = (\beta_{0,2},\ \beta_{1,2},\ \beta_{2,2})'$ are (3×1) vectors of parameters, including the constant. In terms of the S.U.R. representation (21), the above equations can be written as

$$Y_* = ZB + E,$$

where Y_* is a $(T \times 2)$ matrix of observations on vectors y_{1*} and y_{2*}, E is a $(T \times 2)$ matrix whose rows are vectors of normally distributed innovations with variance-covariance $\Sigma = [(\sigma_{\mu\mu'})_{\mu,\mu'=1,2}]$, B is a (3×2)-dimension matrix whose columns, β_1 and β_2, are vectors of parameters and Z is a $(T \times 6)$ matrix whose columns are vectors of possibly collinear variables defined as

$$z_{t1} \equiv z_{t6} \equiv (1-\rho^2)^{1/2} \qquad (t=1),$$
$$z_{t1} \equiv z_{t6} \equiv (1-\rho) \qquad (t=2,3,...,T),$$
$$z_{tj} = \alpha^{1/2}\zeta_{t1} + (1-\alpha)^{1/2}\zeta_{tj} \quad (j=2,3,4,5),$$

where ζ_{tj} $(j = 2,3,4,5)$ are $\mathcal{N}(0,1)$ random variables and α stands for the common correlation coefficient between any two non-constant columns of Z (see also McDonald and Galarneau (1975)). This captures the same degree of multicollinearity between regressors $x_{t1,\mu}$ and $x_{t2,\mu}$ of S.U.R. model (65). In our simulation exercise, we consider the following two values of the collinearity coefficient: $\alpha = 0.5, 0.9$. According to (25), submatrices X_{1*} and X_{2*} (collected in matrix X_*) can be obtained from Z by assuming that submatrices Ψ_1 and Ψ_2, of the block diagonal matrix Ψ are given as follows:

$$\Psi_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}; \quad \Psi_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

In all iterations of our simulation study, the two equations of S.U.R. model (65) were estimated by LS. The residuals of these equations were used to compute the LS estimates of autoregressive coefficients ρ_1 and ρ_2, denoted as $\tilde{\rho}_1$ and $\tilde{\rho}_2$. Then, the transformed variables $y^*_{1,\mu}$ and $x^*_{tj,\mu}$, for $j = 0,1,2$ (where '0' stands for the constant), are calculated as follows:

$$\begin{aligned} y^*_{1,\mu} &= (1-\tilde{\rho}^2_\mu)^{1/2} y_{1,\mu} \quad & x^*_{1j,\mu} &= (1-\tilde{\rho}^2_\mu)^{1/2} x_{1j,\mu} \quad & (t=1), \\ y^*_{t,\mu} &= y_{t,\mu} - \tilde{\rho}_\mu y_{(t-1),\mu} \quad & x^*_{tj,\mu} &= x_{tj,\mu} - \tilde{\rho}_\mu x_{(t-1)j,\mu} \quad & (t \neq 1). \end{aligned} \tag{66}$$

These variables were then used to compute the feasible GL estimates of $\beta_{j,\mu}$ $(j = 0, 1, 2;\ \mu = 1, 2)$, denoted as $\hat{\beta}_{j,\mu}$. The columns of matrix Z were obtained as $z_1 = x^*_{0,1}$, $z_2 = x^*_{1,1}$, $z_3 = x^*_{2,1}$, $z_6 = x^*_{0,2}$, $z_4 = x^*_{1,2}$, $z_5 = x^*_{2,2}$, while the unrestricted estimates of matrix B were based on the GL estimates $\hat{\beta}_{j,\mu}$. The unrestricted estimates of the inverse covariance matrix Σ^{-1} were estimated based on (32) and the feasible GL estimate $\hat{\sigma}_{GL}$ which is calculated by using the following formula:

$$\hat{\sigma}_{GL} = \left[(y - X\hat{\beta})' \left(\hat{P}_I'^{-1}(\hat{\Sigma}_I^{-1} \otimes I_T)\hat{P}_I^{-1}\right)(y - X\hat{\beta})/(MT - n)\right]^{1/2},$$

where I denotes any consistent estimators of matrices Σ^{-1} and P^{-1} (see Appendix), used to obtain a feasible GL estimator of β.

The results of our simlation exercise are presented in Tables 1a, 1b and 2. The actual sizes of our size corrected tests of the following null hypothesis:

$$H_0 : \beta_{2,1} = 0, \tag{67}$$

against its one-sized alternatives, are reported in Tables 1a and 1b. In particular, Table 1a presents results against alternative $H_a : \beta_{2,1} > 0$, while Table 1b against $H_a : \beta_{2,1} < 0$. The table presents the actual sizes (i.e., the rejection probabilities) at the 5% significance level of the following: the standard normal and Student-t tests (denoted as z and t, respectively), their finite-sample size corrected versions based on the Edgeworth corrected critical values of the standard normal and Student-t distributions (denoted as E-z and E-t, respectively) and the Cornish-Fisher finite-sample size corrected Student-t test (denoted as CF-t). Note that we do not examine the performance of the above t tests for the null hypothesis (67) against its two-sided alternatives, since this is a special case of the F test examined in Table 2.

Table 2 presents the actual sizes of our size correceted tests of the following joint null hypothesis on the slope coefficients of S.U.R. model (65), across its two equations:

$$H_0 : \beta_{1,1} = \beta_{2,1} = \beta_{1,2} = \beta_{2,2} = 0. \tag{68}$$

This is done against the alternative hypothesis that at least one of these coefficients are different from zero, i.e., at least one $\beta_{j,\mu} \neq 0$ ($j = 1, 2$; $\mu = 1, 2$). The table presents the actual sizes at the 5% significance level of the following: the standard Wald (chi-square) and F tests (denoted as χ^2 and F, respectively), their finite-sample size corrected versions based on the Edgeworth corrected critical values of the chi-square and F distributions (denoted by E-χ^2 and E-F, respectively) and the Cornish-Fisher finite-sample size corrected F test (denoted as CF-F).

Turning now into the discussion of the results of our simulation study, Tables 1a and 1b clearly indicate that the size corrected tests have better size performance in small samples, like those of $T = 15$ or 20, compared to the standard versions of them based on first order approximations. This is true for both alternatives considered and across all different values of ρ, σ_{12} and a examined.

Between the above different categories of size corrected tests, our results indicate that the CF-t test outperforms the E-z and E-t ones. This is true for almost all cases of a and σ_{12} considered, if ρ takes large values, i.e., $\rho = \pm 0.8$. The same is true when σ_{12} is positive and $\rho = 0.5$. The E-t test outperforms the CF-t test for values of $\rho = \pm 0.5$, when σ_{12} is negative.

Regarding the chi-square and F tests, the results of Table 2 indicate that, in most of the cases examined, the size corrected versions of these tests, i.e., E-χ^2, E-F and CF-F, perform better in small-samples, compared to their standard versions. Between the Edgeworth and Cornish-Fisher size corrected versions of these tests (i.e., *E-F* (or E-χ^2) and *CF-F*), the latter is found to perform better than the former in the case that σ_{12} takes moderate values, i.e., $\sigma_{12} = \pm 0.5$. This is true for all cases of ρ and a considered. On the other hand, the *E-F* and E-χ^2 tests tend to perform better than the *CF-F* test in the case that σ_{12} takes very large values (i.e., $\sigma_{12} = \pm 0.9$), implying a very close to unity correlation coefficient between error terms $u_{t,1}$ and $u_{t,2}$. This, however, happens for moderate values of ρ, i.e., $\rho = \pm 0.5$. For large values of ρ, i.e., $\rho = \pm 0.9$, the *CF-F* test has better size performance than *E-F* (or E-χ^2), even in the case that $\sigma_{12} = \pm 0.9$.

Summing up, the results of our simulation exercise clearly indicate that the finite-sample size corrected tests E-χ^2, *E-F* and *CF-F* can considerably improve the performance of the standard (uncorrected) tests in small-samples. This happens even for very high levels of autocorrelation and/or cross-correlation between the error terms of the equations of the S.U.R. model. Another interesting conclusion that can be drawn from the results of this exercise is that the adjusted for the degrees of freedom versions of the tests perform better than their unadjusted ones in most of the cases of our simulation exercise considered. Note that this is also true for the standard (uncorrected) versions of the tests.

Table 1a: $H_0 : \beta_{2,1} = 0$ against $H_A : \beta_{2,1} > 0$ (Nominal size: 5%)

			Actual sizes (%)																			
Test:			z	E-z	t	E-t	CF-t	z	E-z	t	E-t	CF-t	z	E-z	t	E-t	CF-t	z	E-z	t	E-t	CF-t
α	σ_{12}	T	$\rho = -0.8$					$\rho = -0.5$					$\rho = 0.5$					$\rho = 0.8$				
0.5	-0.90	15	11.2	7.2	10.5	7.2	5.3	7.4	4.7	6.8	4.9	4.3	6.4	3.8	5.8	4.1	3.7	8.2	5.0	7.6	5.2	4.4
		20	10.9	7.4	10.3	7.5	6.1	6.7	4.4	6.3	4.6	4.3	5.7	3.5	5.2	3.7	3.4	7.2	4.5	6.6	4.6	4.0
		40	7.5	5.9	7.4	6.1	5.6	5.5	4.2	5.3	4.4	4.3	4.4	3.3	4.3	3.5	3.4	6.6	4.7	6.4	4.8	4.5
	-0.75	15	10.8	6.8	10.2	7.0	5.1	7.7	4.9	7.0	5.1	4.6	6.8	3.9	6.1	4.2	3.6	8.6	5.6	8.1	5.8	4.9
		20	11.2	7.1	10.6	7.2	5.8	7.2	5.2	6.7	5.4	4.9	6.0	3.8	5.5	4.0	3.7	8.8	5.3	8.2	5.5	4.9
		40	8.1	5.6	7.8	5.8	5.5	6.0	4.7	5.8	4.8	4.8	5.6	4.3	5.4	4.5	4.4	6.6	4.7	6.5	4.9	4.5
	-0.50	15	10.9	6.6	10.2	6.7	4.7	7.3	4.5	6.8	4.7	4.2	7.5	4.8	6.7	5.1	4.7	10.2	6.6	9.6	6.9	5.9
		20	10.1	6.9	9.8	7.1	5.6	6.8	4.4	6.4	4.5	4.4	6.6	4.4	6.3	4.6	4.3	8.8	5.9	8.3	6.0	5.2
		40	7.3	5.5	7.1	5.6	5.1	6.7	5.3	6.5	5.4	5.3	6.0	4.8	5.8	4.9	4.8	7.9	5.8	7.7	6.0	5.6
	0.50	15	9.2	5.9	8.4	6.0	4.0	6.7	4.1	6.1	4.2	3.8	7.5	5.3	6.9	5.4	5.0	11.9	8.1	11.0	8.3	7.0
		20	8.1	5.0	7.7	5.2	4.0	5.5	3.4	5.1	3.7	3.3	7.5	5.2	6.9	5.3	4.9	10.2	7.0	9.7	7.2	6.1
		40	6.1	4.1	5.9	4.3	3.9	5.6	4.4	5.5	4.5	4.5	6.6	5.3	6.4	5.4	5.3	8.7	6.6	8.3	6.8	6.4
	0.75	15	8.3	4.8	7.6	4.8	3.4	6.3	3.7	5.7	3.9	3.4	7.6	4.9	6.8	5.0	4.6	11.8	7.9	10.9	8.0	7.0
		20	7.5	4.4	7.1	4.5	3.7	5.7	3.6	5.3	3.7	3.5	7.3	4.9	6.9	5.1	4.8	10.6	7.4	10.1	7.6	6.8
		40	6.0	4.1	5.7	4.3	4.0	4.7	3.3	4.5	3.5	3.4	5.9	4.4	5.8	4.6	4.5	8.3	6.5	8.1	6.6	6.2
	0.90	15	7.4	4.0	6.7	4.1	3.2	5.5	3.0	5.0	3.1	2.6	8.1	5.0	7.2	5.2	4.8	11.2	7.7	10.5	7.9	6.8
		20	7.0	4.0	6.6	4.2	3.3	5.3	2.9	4.7	3.1	2.9	7.5	5.1	7.0	5.3	5.1	10.2	7.2	9.8	7.4	6.1
		40	5.2	3.4	5.0	3.6	3.3	4.5	3.3	4.3	3.5	3.4	5.7	4.4	5.4	4.4	4.3	8.1	6.0	8.0	6.2	5.9
0.9	-0.90	15	12.3	7.6	11.5	7.7	5.5	7.9	5.1	7.3	5.3	4.7	6.5	3.9	5.9	4.2	3.8	8.9	5.5	8.0	5.8	4.8
		20	11.5	7.3	10.9	7.3	5.9	6.6	4.4	6.1	4.6	4.4	6.3	4.4	5.8	4.7	4.3	7.8	4.9	7.3	5.1	4.4
		40	7.7	5.6	7.5	5.8	5.3	5.8	4.5	5.6	4.6	4.5	4.8	3.5	4.6	3.6	3.6	5.6	4.2	5.4	4.2	4.0
	-0.75	15	12.5	7.7	11.7	7.8	5.4	7.4	4.5	6.7	4.7	4.1	7.2	4.7	6.5	4.9	4.5	9.5	6.1	8.7	6.3	5.3
		20	11.3	7.8	10.7	7.9	6.1	8.0	5.4	7.3	5.6	5.3	6.4	4.4	6.1	4.6	4.2	8.8	5.7	8.3	5.8	4.9
		40	7.7	5.6	7.4	5.9	5.4	6.2	4.9	6.0	5.1	5.0	5.7	4.1	5.5	4.3	4.1	7.0	5.1	6.9	5.2	4.9
	-0.50	15	11.1	6.9	10.2	6.9	5.1	7.4	4.3	6.6	4.5	4.0	7.7	4.8	7.0	5.1	4.6	9.3	6.0	8.7	6.2	5.3
		20	10.6	7.4	10.1	7.5	5.8	7.8	5.1	7.3	5.4	5.0	6.5	4.3	6.1	4.5	4.2	8.5	5.8	8.1	6.0	5.3
		40	8.0	5.7	7.9	5.9	5.6	6.2	4.7	6.1	4.9	4.8	6.2	4.6	6.0	4.7	4.6	6.8	4.9	6.7	5.1	4.7
	0.50	15	8.7	5.4	8.0	5.5	3.9	7.2	4.5	6.5	4.8	4.1	8.6	5.6	7.9	5.8	5.3	12.1	8.3	11.3	8.5	7.2
		20	8.7	5.6	8.3	5.8	4.4	6.0	4.0	5.5	4.2	3.9	7.8	5.3	7.2	5.6	5.1	10.9	7.6	10.3	7.9	6.8
		40	5.7	4.1	5.6	4.2	3.9	5.2	4.0	5.0	4.0	4.0	5.9	4.6	5.7	4.8	4.7	8.1	5.7	7.8	5.7	5.5
	0.75	15	8.4	5.1	7.8	5.3	3.7	6.3	3.8	5.7	4.0	3.5	8.4	5.3	7.7	5.6	5.1	11.6	8.2	11.0	8.4	7.1
		20	8.7	5.0	8.2	5.2	4.0	6.0	3.6	5.4	3.8	3.5	7.8	5.2	7.5	5.4	5.1	10.9	7.4	10.3	7.6	6.7
		40	6.4	4.2	6.1	4.3	4.1	5.3	4.1	5.0	4.2	4.1	6.7	5.4	6.6	5.5	5.4	8.6	6.7	8.3	6.8	6.5
	0.90	15	7.3	4.3	6.8	4.4	3.3	5.3	3.0	4.8	3.2	2.8	8.0	5.2	7.5	5.5	4.9	11.8	7.7	11.0	7.9	6.8
		20	7.4	4.3	7.0	4.5	3.5	5.5	3.4	5.1	3.6	3.3	7.6	5.4	7.0	5.6	5.2	10.6	7.3	10.0	7.4	6.6
		40	6.3	4.5	6.0	4.7	4.5	4.9	3.6	4.7	3.7	3.6	6.4	4.9	6.0	5.0	4.9	8.1	6.3	7.8	6.4	6.2

Table 1b: $H_0 : \beta_{2,1} = 0$ against $H_A : \beta_{2,1} < 0$ (Nominal size: 5%)

			Actual sizes (%)																			
Test:			z	E-z	t	E-t	CF-t	z	E-z	t	E-t	CF-t	z	E-z	t	E-t	CF-t	z	E-z	t	E-t	CF-t
α	σ_{12}	T	$\rho = -0.8$					$\rho = -0.5$					$\rho = 0.5$					$\rho = 0.8$				
0.5	-0.90	15	11.2	6.7	10.2	6.8	5.1	7.0	4.4	6.4	4.6	3.9	6.7	3.4	6.0	3.9	3.3	8.8	5.4	8.1	5.5	4.8
		20	10.2	6.5	9.8	6.6	5.1	7.0	4.5	6.6	4.8	4.3	5.7	3.6	5.3	3.8	3.4	8.7	5.3	8.1	5.6	4.7
		40	6.9	5.2	6.7	5.4	4.9	5.1	3.9	4.9	4.1	4.0	5.0	3.7	4.8	3.8	3.7	5.9	4.3	5.7	4.4	4.1
	-0.75	15	11.0	6.6	10.2	6.7	4.7	8.0	5.2	7.4	5.5	4.9	6.4	4.2	5.8	4.4	4.1	9.1	6.0	8.5	6.2	5.3
		20	10.2	6.4	9.7	6.5	5.2	6.8	4.6	6.3	4.8	4.5	6.3	4.0	5.8	4.2	3.9	8.5	5.5	7.9	5.7	4.8
		40	7.7	5.4	7.3	5.6	5.2	5.7	4.5	5.5	4.7	4.6	4.4	3.3	4.2	3.4	3.3	7.0	5.0	6.8	5.2	4.8
	-0.50	15	10.7	6.7	9.8	6.8	5.1	7.8	4.7	7.0	4.9	4.2	7.5	4.8	6.8	5.0	4.5	9.5	6.2	8.7	6.4	5.4
		20	10.1	6.4	9.5	6.5	5.0	7.2	5.0	6.7	5.3	5.0	7.1	4.4	6.6	4.7	4.3	8.6	5.5	7.9	5.6	4.9
		40	7.7	5.7	7.6	5.8	5.4	6.3	4.9	6.1	5.1	4.9	5.8	4.7	5.6	4.9	4.7	7.0	5.4	6.8	5.6	5.3
	0.50	15	9.4	5.4	8.5	5.6	4.0	6.2	3.9	5.4	4.0	3.8	9.2	6.1	8.6	6.3	5.9	10.9	7.6	10.1	7.7	6.5
		20	7.5	4.5	7.1	4.7	3.5	6.6	3.8	6.2	4.1	3.7	7.5	4.9	6.9	5.1	4.9	11.1	7.7	10.6	8.0	6.8
		40	6.0	4.7	5.9	4.7	4.5	5.6	3.9	5.4	4.1	4.0	6.9	5.3	6.5	5.4	5.3	7.9	6.0	7.6	6.1	5.8
	0.75	15	8.0	4.9	7.4	4.9	3.7	5.6	3.3	5.0	3.5	3.1	7.9	5.2	7.1	5.4	4.9	11.8	8.2	11.1	8.4	7.3
		20	8.0	4.7	7.4	4.9	4.0	5.2	3.3	4.6	3.4	3.1	7.7	5.1	7.2	5.3	5.0	10.5	7.6	9.9	7.7	6.9
		40	5.7	3.8	5.5	4.0	3.6	5.4	4.1	5.3	4.2	4.2	5.4	4.3	5.3	4.4	4.4	8.2	6.0	7.8	6.1	5.9
	0.90	15	6.9	4.1	6.3	4.2	3.1	5.0	2.5	4.3	2.6	2.3	7.5	4.8	6.8	4.9	4.5	11.3	7.5	10.5	7.7	6.6
		20	6.4	3.5	5.9	3.6	2.6	5.2	2.9	4.7	3.2	2.8	6.8	4.6	6.5	4.9	4.4	11.2	7.9	10.5	8.1	7.3
		40	4.9	2.9	4.7	3.1	2.8	4.4	3.1	4.2	3.2	3.1	5.4	4.1	5.3	4.2	4.1	7.4	5.5	7.2	5.5	5.2
0.9	-0.90	15	12.4	8.1	11.3	8.2	5.8	7.8	4.8	7.0	5.0	4.4	6.3	3.8	5.6	4.0	3.5	8.8	5.5	8.0	5.5	4.6
		20	10.8	7.0	10.2	7.2	5.7	7.0	4.6	6.5	4.8	4.2	6.4	3.9	5.7	4.1	3.7	8.4	5.6	7.9	5.8	4.6
		40	7.3	5.4	7.1	5.6	5.2	5.6	4.3	5.4	4.5	4.4	5.0	3.8	4.8	4.0	3.9	6.4	4.5	6.0	4.5	4.4
	-0.75	15	11.9	7.2	11.0	7.4	5.3	7.0	4.5	6.1	4.6	4.2	7.0	4.6	6.5	4.9	4.4	8.8	5.6	8.1	5.7	4.8
		20	9.7	6.6	9.2	6.7	5.4	7.8	5.0	7.2	5.3	4.9	6.8	4.9	6.4	5.2	4.7	8.4	5.4	7.9	5.5	4.9
		40	8.7	6.3	8.5	6.4	6.2	5.9	4.6	5.8	4.8	4.7	4.9	4.1	4.9	4.1	4.1	7.1	5.3	6.9	5.5	5.0
	-0.50	15	11.2	7.0	10.1	7.1	5.1	8.1	4.8	7.3	5.0	4.5	6.8	4.5	6.2	4.6	4.3	10.1	6.2	9.2	6.5	5.5
		20	9.8	6.1	9.4	6.2	4.8	6.7	4.6	6.3	4.7	4.5	6.9	4.5	6.4	4.6	4.4	9.0	6.1	8.6	6.4	5.5
		40	7.5	5.4	7.4	5.7	5.3	6.7	5.2	6.4	5.4	5.3	5.5	4.1	5.2	4.2	4.2	7.8	5.6	7.6	5.7	5.5
	0.50	15	9.5	5.7	8.8	5.7	4.2	6.3	3.7	5.7	3.9	3.6	8.1	5.7	7.4	5.9	5.5	11.9	8.3	11.3	8.5	7.1
		20	8.1	5.0	7.6	5.2	4.1	6.9	4.3	6.4	4.6	4.3	7.5	5.1	7.0	5.3	5.0	11.0	7.7	10.5	7.8	6.8
		40	6.4	4.5	6.3	4.7	4.4	5.1	4.0	4.9	4.2	4.1	6.3	5.2	6.1	5.3	5.2	8.7	6.3	8.5	6.4	6.1
	0.75	15	8.6	5.0	8.0	5.0	3.6	5.9	3.2	5.3	3.4	3.0	8.5	5.5	7.7	5.7	5.2	12.0	8.3	11.2	8.5	7.3
		20	7.6	4.2	7.0	4.3	3.2	5.7	3.4	5.3	3.6	3.3	7.9	5.6	7.5	5.8	5.4	10.7	7.4	10.2	7.5	6.8
		40	6.1	4.5	5.9	4.6	4.2	4.4	3.3	4.2	3.5	3.4	6.9	5.3	6.6	5.4	5.3	8.9	6.6	8.6	6.7	6.4
	0.90	15	8.4	4.8	7.8	4.8	3.6	5.1	3.0	4.4	3.1	2.8	7.9	5.4	7.2	5.6	5.2	11.1	7.6	10.4	7.8	6.6
		20	7.1	4.3	6.5	4.5	3.5	5.6	3.5	5.2	3.8	3.3	7.7	5.5	7.3	5.7	5.4	11.3	8.0	10.7	8.2	7.3
		40	5.7	3.3	5.4	3.5	3.1	4.5	3.1	4.2	3.2	3.2	5.4	4.4	5.2	4.5	4.4	8.6	6.6	8.4	6.8	6.5

Table 2: $H_0 : \beta_{1,1} = \beta_{2,1} = \beta_{1,2} = \beta_{2,2} = 0$ (Nominal size: 5%)

Test:			Actual sizes (%)																			
			χ^2	E-χ^2	F	E-F	CF-F	χ^2	E-χ^2	F	E-F	CF-F	χ^2	E-χ^2	F	E-F	CF-F	χ^2	E-χ^2	F	E-F	CF-F
α	σ_{12}	T	$\rho = -0.8$					$\rho = -0.5$					$\rho = 0.5$					$\rho = 0.8$				
0.5	-0.90	15	30.5	17.0	24.6	14.1	2.8	12.6	5.2	8.6	4.6	2.7	11.4	5.7	8.4	5.3	3.8	21.7	12.0	17.0	10.6	5.3
		20	25.5	14.3	21.4	13.0	3.7	11.0	5.1	8.4	4.9	3.7	9.1	4.2	7.1	4.0	2.9	16.6	8.9	13.7	8.4	4.3
		40	13.4	7.7	11.7	7.8	5.6	6.9	3.9	5.9	4.1	3.8	4.5	2.3	3.9	2.4	2.1	8.8	4.7	7.6	4.8	3.4
	-0.75	15	32.3	18.5	26.7	15.3	2.5	14.4	6.6	10.0	6.1	3.8	14.7	7.4	10.5	7.0	5.2	25.2	15.9	20.8	14.2	7.4
		20	27.7	16.2	23.7	14.8	4.2	12.4	6.2	9.7	6.2	4.6	11.1	5.7	8.8	5.6	4.4	21.0	12.5	17.8	11.7	6.1
		40	13.5	8.2	12.3	8.3	6.5	6.5	4.0	5.6	4.2	3.9	6.4	3.4	5.4	3.7	3.3	11.3	6.7	10.2	6.8	5.5
	-0.50	15	32.9	19.6	27.4	16.8	3.1	16.4	8.0	11.7	7.4	4.9	16.9	9.1	12.6	8.6	6.2	30.0	18.5	24.5	16.8	9.1
		20	27.9	15.6	23.7	14.5	4.5	13.2	6.1	10.3	6.3	4.9	13.0	6.9	9.9	7.0	5.6	23.5	14.2	20.2	13.4	7.3
		40	15.7	9.7	14.1	9.8	7.6	8.2	4.6	7.1	5.1	4.6	8.0	4.2	6.6	4.5	4.2	13.1	7.9	12.0	8.1	6.3
	0.50	15	27.0	16.0	22.0	13.5	2.5	13.4	6.3	9.5	5.8	4.1	18.6	9.7	14.3	9.1	6.9	33.3	20.2	26.6	18.5	9.8
		20	21.7	12.6	18.2	11.6	3.8	9.8	5.2	7.8	5.2	4.1	14.5	7.9	11.4	7.9	6.5	30.8	18.9	26.1	18.0	10.0
		40	11.0	6.3	9.7	6.3	4.7	6.4	3.4	5.4	3.7	3.4	9.2	5.3	7.8	5.7	5.2	18.1	11.4	16.2	11.5	9.4
	0.75	15	22.5	12.4	18.2	10.5	2.0	10.4	4.8	7.2	4.4	2.9	16.2	8.0	11.9	7.6	5.8	31.0	19.1	25.7	16.9	9.3
		20	17.5	9.8	14.5	9.1	2.7	8.6	3.8	6.3	3.8	2.9	13.2	7.1	10.2	7.0	5.4	29.6	17.9	25.5	16.9	9.2
		40	9.3	4.9	7.9	5.0	3.6	5.0	2.8	4.2	2.9	2.7	7.8	5.0	6.7	5.2	4.8	15.3	9.4	13.6	9.6	7.7
	0.90	15	18.2	10.1	14.3	8.6	1.5	8.1	3.3	5.3	2.8	1.6	14.7	6.7	10.7	6.1	4.6	28.6	16.1	22.7	14.1	7.1
		20	14.8	7.8	12.3	6.9	2.0	6.4	3.0	4.8	2.9	2.1	12.1	6.1	9.7	6.1	4.3	26.2	14.9	21.5	13.9	7.3
		40	7.9	4.0	7.0	4.1	2.8	3.9	2.0	3.3	2.2	2.0	6.5	3.2	5.5	3.6	3.1	15.1	9.3	13.5	9.3	7.2
0.9	-0.90	15	25.0	13.4	19.9	11.4	1.8	9.3	4.2	6.2	3.8	2.4	7.9	3.7	5.4	3.5	2.5	16.7	9.5	13.4	8.5	4.7
		20	19.9	11.4	16.8	10.4	3.5	7.5	3.2	5.4	3.2	2.4	6.3	3.1	4.9	3.0	2.3	13.2	7.6	11.2	7.1	3.5
		40	9.5	5.3	8.5	5.4	4.0	3.8	0.2	3.2	2.3	2.1	2.7	1.8	2.5	1.8	1.7	6.0	3.1	5.6	3.3	2.2
	-0.75	15	29.8	16.5	24.2	14.0	2.6	11.4	5.3	8.2	4.9	3.4	12.3	6.8	9.1	6.5	5.0	22.3	13.8	18.0	12.7	6.6
		20	24.5	13.4	20.2	12.5	3.8	10.2	4.6	7.4	4.6	3.5	8.8	4.5	6.8	4.4	3.6	18.7	10.6	15.6	10.1	5.0
		40	10.2	5.9	9.1	6.0	4.4	5.1	2.9	4.4	3.4	2.9	3.9	2.2	3.3	2.3	2.2	8.9	5.0	7.8	5.2	3.8
	-0.50	15	32.0	18.3	26.0	15.7	2.7	14.9	7.4	10.8	6.9	5.1	16.5	8.8	12.2	8.2	6.5	28.5	17.5	23.4	16.2	8.3
		20	26.5	15.0	22.7	13.3	4.2	12.6	6.2	9.9	6.2	4.8	12.1	6.3	9.6	6.3	5.0	23.2	14.0	19.3	13.5	7.7
		40	13.0	7.6	11.6	7.7	5.8	7.6	4.0	6.2	4.3	4.0	7.4	3.9	6.4	4.3	3.8	11.8	6.7	10.4	6.9	5.1
	0.50	15	25.1	14.6	20.3	12.7	2.1	13.8	6.7	9.9	6.3	4.4	17.7	9.4	13.4	8.8	6.9	33.7	21.3	27.5	19.4	10.0
		20	21.6	12.1	18.4	11.1	3.1	9.2	4.5	7.1	4.5	3.6	13.5	7.3	10.5	7.3	5.8	28.8	18.0	24.9	17.0	9.2
		40	9.8	5.3	8.5	5.4	4.1	5.8	3.2	4.7	3.4	3.2	8.4	5.2	7.4	5.4	4.9	15.9	9.6	14.3	9.8	8.0
	0.75	15	19.9	11.5	16.1	9.9	1.7	8.6	3.9	6.0	3.7	2.5	13.7	6.9	10.0	6.4	5.0	29.7	17.5	23.8	16.0	8.6
		20	15.9	9.1	13.4	8.5	2.6	6.7	2.9	4.8	2.9	2.4	11.1	5.7	8.8	5.8	4.5	25.5	14.9	21.2	14.3	7.7
		40	7.1	4.0	6.4	4.0	2.9	4.0	2.2	3.3	2.4	2.1	5.0	2.5	4.2	2.7	2.4	12.6	7.6	11.2	7.8	5.8
	0.90	15	15.1	8.5	11.9	7.1	1.5	6.5	2.7	4.4	2.4	1.7	10.1	4.4	7.3	4.0	2.9	22.2	12.2	17.2	10.9	5.6
		20	11.7	6.5	9.6	5.9	1.5	4.9	2.1	3.7	2.0	1.5	8.5	4.0	6.3	3.9	3.1	21.1	12.0	17.3	11.1	6.3
		40	5.3	2.8	4.7	2.8	2.0	2.4	1.2	2.0	1.3	1.2	3.9	2.3	3.5	2.5	2.3	10.6	6.4	9.4	6.4	5.2

6 Conclusions

In this paper, we have employed Edgeworth expansions of the standard normal (or Student-t) and chi-square (or F) distributions to derive second-order size corrected testing procedures for the coefficient of the S.U.R. model with first-order autocorrelated errors. These procedures include (i) the Edgeworth corrected critical values of the well-known Wald (or F) and t tests and (ii) the Cornish-Fisher corrected F and t test statistics. Since the standard F and t tests are adjusted for the degrees of freedom, they are locally exact, which means that their approximate distributions become exact when the model is sufficiently simplified.

The Edgeworth and Cornish-Fisher expansions, employed by the paper, are equivalent to each other, since the latter constitutes an inversion of the former. However, in practice, the use of the Cornish-Fisher corrected test statistics is recommended, since they are proper random variables with well-behaved distribution tails. The Edgeworth approximation can assign negative 'probabilities' in the tails of the approximate distributions. Furthermore, the Cornish-Fisher size corrected tests can be easily implemented, in practice, using the tables of the Student-t and the F distributions.

To evaluate the small-sample performance of the suggested tests, the paper has conducted a Monte Carlo study. The results of this exercise indicate that the size corrected t and F tests lead to substantial size improvements upon their standard versions assuming first-order asymptotic approximations. This is true even for very small samples of 15 or 20 observations. Between the Edgeworth and Cornish-Fisher categories of the size corrected tests suggested in the paper, the second category is found to perform better than the first, for most cases of serial and cross-equation correlation of the error terms of the S.U.R. model examined. This result is also robust across different degrees of multicollinearity between the independent variables of the model considered. In particular, both the t and F Cornish-Fisher size corrected tests are found to outperform their Edgeworth size corrected counterparts, when the degree of serial correlation of the error terms is very high, as often observed in practice. For the t test, this is true even for a close-to-unity degree of correlation across the two equations of the S.U.R. model.

Appendix

In this appendix, we provide proofs of the main results of the paper. To prove these results, we rely on a number of lemmas. Some of them are given without proof for reasons of space. These proofs are available upon request. The presentation of our proofs is scheduled as follows: First, we provide some preliminary matrix-algebra results, needed for the calculation of the quantities in the stochastic expansions of all estimators considered and the tests. Then, given these lemmas, we give the proofs of the theorems.

Matrix-algebra results

Following Magdalinos (1992, page 344), let $\mathcal{I}$ be a given set of indices which, without loss of generality, can be considered to belong to the open interval $(0, 1)$. For any collection of real-valued stochastic quantities (scalars, vectors, or matrices) Y_τ ($\tau \in \mathcal{I}$), we write $Y_\tau = \omega(\tau^i)$, if for any given $n > 0$, there exists a $0 < \epsilon < \infty$ such that

$$\Pr\left[\|Y_\tau/\tau^i\| > (-\ln\tau)^\epsilon\right] = o(\tau^n), \tag{A.1}$$

as $\tau \to 0$, where the $\|\cdot\|$ is the Euclidean norm. If (A.1) is valid for any $n > 0$, we write $Y_\tau = \tau(\infty)$. The use of this order of magnitude is motivated by the fact that, if two stochastic quantities differ by a quantity of order $\omega(\tau^i)$, then, under general conditions, the distribution function of the one provides an asymptotic approximation of the distribution function of the other, with an error of order $O(\tau^i)$. Furthermore, orders $\omega(\cdot)$ and $O(\cdot)$ have similar operational properties (Magdalinos (1992)).

Define the following $(T \times T)$ matrices: D whose (t,t')-th element is equal to 1 if $|t-t'| = 1$ and 0 elsewhere, D_j whose (t,t')-th element is equal to 1 if $t-t' = 1$ and 0 elsewhere, D_i whose (t,t')-th element is equal to 1 if $t-t' = -1$ and 0 elsewhere. Also, define the following $(T \times T)$ matrices: Δ with 1 in $(1,1)$-st and (T,T)-th positions and 0's elsewhere, Δ_{11} with 1 in $(1,1)$-st position and 0's elsewhere, Δ_{TT} with 1 in (T,T)-th position and 0's elsewhere. Moreover, by using matrix P_μ in (9), we can calculate $(T \times T)$ matrices R_{ij} as follows:

$$R_{ij} = P_i P_j' = \frac{1}{1-\rho_i\rho_j}\begin{bmatrix} 1 & \rho_j & \cdots & \rho_j^{T-1} \\ \rho_i & 1 & \cdots & \rho_j^{T-2} \\ \vdots & \vdots & & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \cdots & 1 \end{bmatrix}. \qquad \text{(A.2)}$$

Matrices R^{ij} help us to write the elements of matrix Ω analytically. For these matrices and their derivatives the following two lemmas hold:

Lemma A.1. *For matrix R^{ii}, which is the inverse of R_{ii}, the following result holds:*

$$R^{ii} = P_i'^{-1} P_i^{-1} = (1+\rho_i^2) I_T - \rho_i D - \rho_i^2 \Delta, \qquad \text{(A.3)}$$

where $R^{ii} = R_{ii}^{-1}$ $(\forall i)$. Moreover, for matrix R^{ij}, the following result holds:

$$\begin{aligned} R^{ij} = P_i'^{-1} P_j^{-1} &= (1+\rho_i\rho_j) I_T - \rho_i D_i - \rho_j D_j - \rho_i\rho_j \Delta_{TT} \\ &\quad + [(1-\rho_i^2)^{1/2}(1-\rho_j^2)^{1/2} - 1]\Delta_{11}. \end{aligned} \qquad \text{(A.4)}$$

Note that R^{ij} is not the inverse of R_{ij}, i.e., $R^{ij} \neq R_{ij}^{-1}$ $(\forall i \neq j)$.

Proof of Lemma A.1. The proof is straightforward. □

Define the $(M \times M)$ matrix $\Sigma^{-1} = [(\sigma^{\mu\mu'})_{\mu,\mu'=1,\,\dots,\,M}]$ and scalars:

$$\alpha_{ij} = (1-\rho_i^2)^{1/2}(1-\rho_j^2)^{1/2},$$
$$\xi'_{(i)j} = \partial\alpha_{ij}/\partial\rho_i, \quad \xi''_{(i)j} = \partial^2\alpha_{ij}/\partial^2\rho_i, \quad \xi''_{(i)(j)} = \partial^2\alpha_{ij}/\partial\rho_i\partial\rho_j, \tag{A.5}$$
$$R^{ij}_{\rho_\mu} = \partial R^{ij}/\partial\rho_\mu, \quad R^{ij}_{\rho_\mu\rho_{\mu'}} = \partial^2 R^{ij}/\partial\rho_\mu\partial\rho_{\mu'}.$$

Lemma A.2. *The following results on the partial derivatives of matrix R^{ij} hold:*

$$R^{ii}_{\rho_i} = 2\rho_i I_T - D - 2\rho_i\Delta, \quad R^{ii}_{\rho_i\rho_i} = 2(I_T-\Delta) \quad (\forall i),$$
$$R^{ii}_{\rho_j} = R^{ii}_{\rho_j\rho_j} = R^{ii}_{\rho_i\rho_j} = 0 \quad (\forall i \neq j),$$
$$R^{ij}_{\rho_i} = \rho_j I_T - D_i - \rho_j\Delta_{TT} + \xi'_{(i)j}\Delta_{11} \quad (\forall i,j), \tag{A.6}$$
$$R^{ij}_{\rho_i\rho_i} = \xi''_{(i)j}\Delta_{11}, \quad R^{ij}_{\rho_i\rho_j} = I_T - \Delta_{TT} + \xi''_{(i)(j)}\Delta_{11} \quad (\forall i,j),$$
$$R^{ij}_{\rho_\mu} = R^{ij}_{\rho_\mu\rho_\mu} = R^{ij}_{\rho_\mu\rho_i} = R^{ij}_{\rho_\mu\rho_j} = 0 \quad (\forall\mu \neq i \wedge \forall\mu \neq j),$$

with obvious modifications for $R^{ij}_{\rho_j}$ and $R^{ij}_{\rho_j\rho_j}$. Further,

$$\begin{aligned}
\xi'_{(i)j} &= -\rho_i(1-\rho_i^2)^{-1/2}(1-\rho_j^2)^{1/2} \quad (\forall i),\\
\xi''_{(i)j} &= -(1-\rho_i^2)^{-3/2}(1-\rho_j^2)^{1/2} \quad (\forall i),\\
\xi''_{(i)(j)} &= \rho_i\rho_j(1-\rho_i^2)^{-1/2}(1-\rho_j^2)^{-1/2} \quad (\forall i,j),
\end{aligned} \tag{A.7}$$
$$\frac{\partial\alpha_{ij}}{\partial\rho_\mu} = \frac{\partial^2\alpha_{ij}}{\partial\rho_\mu^2} = \frac{\partial^2\alpha_{ij}}{\partial\rho_\mu\partial\rho_i} = \frac{\partial^2\alpha_{ij}}{\partial\rho_\mu\partial\rho_j} = 0 \quad (\forall\mu \neq i \wedge \forall\mu \neq j).$$

Proof of Lemma A.2. The proof follows using Lemma A.1, after tedious algebra.

□

Lemma A.3. *For the elements of matrix* Ω *can be analytically written as follows:*

$$\begin{aligned}
\sum\nolimits_{k=1}^{M} \sigma_{ik}\sigma^{ki} &= \sum\nolimits_{k=1}^{M} \sigma^{ik}\sigma_{ki} = 1, \\
\sum\nolimits_{k=1}^{M} \sigma_{ik}\sigma^{kj} &= \sum\nolimits_{k=1}^{M} \sigma^{ik}\sigma_{kj} = 0 \ \ (\forall i \neq j), \\
\sum\nolimits_{k=1}^{M} \sigma_{ik}\sigma^{ki} R_{ik}R^{ki} &= \sum\nolimits_{k=1}^{M} \sigma^{ik}\sigma_{ki} R^{ik}R_{ki} = I_{TM}, \\
\sum\nolimits_{k=1}^{M} \sigma_{ik}\sigma^{kj} R_{ik}R^{kj} &= \sum\nolimits_{k=1}^{M} \sigma^{ik}\sigma_{kj} R^{ik}R_{kj} = 0 \ \ (\forall i \neq j).
\end{aligned} \tag{A.8}$$

Proof of Lemma A.3. The results of the lemma can be proved by noticing that that

$$\Omega^{-1} = P(\Sigma \otimes I_T)P' = [(\sigma_{ij}R_{ij})_{i,j=1,\,\ldots,\,M}] \Rightarrow \Omega = [(\sigma^{ij}R^{ij})_{i,j=1,\,\ldots,\,M}], \tag{A.9}$$

since P is block diagonal, $\Sigma\Sigma^{-1} = \Sigma^{-1}\Sigma = I_M$ and $\Omega\Omega^{-1} = \Omega^{-1}\Omega = I_{TM}$. □

To derive the partial derivatives of Ω with respect to nuisance parameters, given in the next lemma, we need the following definitions. For the composite index $(ij) = 1, \ldots, M^2$, defined in (20), let $\varsigma_{(ij)} = \sigma^{ij}$ be the elements of the $(M^2 \times 1)$ vector $\varsigma = vec(\Sigma^{-1})$. Also, let $\Delta_{\mu\mu'} = [(\delta_{\mu i}\delta_{j\mu'})_{i,j=1,\,,\ldots,\,M}]$ be a $(M \times M)$ matrix with 1 in the (μ, μ')-th position and 0's elsewhere. Then, for all μ, μ', ν and ν', we have

$$\frac{\partial}{\partial\varsigma_{(\mu\mu')}}(\Sigma^{-1} \otimes I_T) = \Delta_{\mu\mu'} \otimes I_T, \quad \frac{\partial^2}{\partial\varsigma_{(\mu\mu')}\partial\varsigma_{(\nu\nu')}}(\Sigma^{-1} \otimes I_T) = 0. \tag{A.10}$$

Lemma A.4. *The partial derivatives of* Ω *with respect to the elements of vectors* ϱ *and* ς*, can be analytically written as follows:*

$$\Omega_{\varsigma_{(\mu\mu')}} = [(\delta_{\mu i}\delta_{j\mu'}R^{\mu\mu'})_{i,j=1,\,\ldots,\,M}], \quad \Omega_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}} = 0, \tag{A.11}$$

$$\begin{aligned}
\Omega_{\rho_\mu} &= [(\delta_{\mu i}\sigma^{\mu j}R^{\mu j}_{\rho_\mu} + \delta_{j\mu}\sigma^{i\mu}R^{i\mu}_{\rho_\mu} + \delta_{\mu i}\delta_{j\mu}\sigma^{\mu\mu}R^{\mu\mu}_{\rho_\mu})_{i,j=1,\,\ldots,\,M}], \\
\Omega_{\rho_\mu\rho_\mu} &= [(\delta_{\mu i}\sigma^{\mu j}R^{\mu j}_{\rho_\mu\rho_\mu} + \delta_{j\mu}\sigma^{i\mu}R^{i\mu}_{\rho_\mu\rho_\mu} + \delta_{\mu i}\delta_{j\mu}\sigma^{\mu\mu}R^{\mu\mu}_{\rho_\mu\rho_\mu})_{i,j=1,\,\ldots,\,M}], \\
\Omega_{\rho_\mu\rho_{\mu'}} &= [(\delta_{\mu i}\delta_{j\mu'}\sigma^{\mu\mu'}R^{\mu\mu'}_{\rho_\mu\rho_{\mu'}} + \delta_{\mu' i}\delta_{j\mu}\sigma^{\mu'\mu}R^{\mu'\mu}_{\rho_\mu\rho_{\mu'}})_{i,j=1,\,\ldots,\,M}],
\end{aligned} \tag{A.12}$$

$$\Omega_{\rho_\mu \varsigma_{(\nu\nu')}} = [(\delta_{\nu i}\delta_{j\nu'}\delta_{\mu\nu}R^{\mu\nu'}_{\rho_\mu} + \delta_{\nu i}\delta_{j\nu'}\delta_{\nu'\mu}R^{\nu\mu}_{\rho_\mu} + \delta_{\nu i}\delta_{j\nu'}\delta_{\mu\nu}\delta_{\nu'\mu}R^{\mu\mu}_{\rho_\mu})_{i,j=1,\,\dots,\,M}] \tag{A.13}$$

$$\Rightarrow \Omega_{\rho_\mu \varsigma_{(\nu\nu')}} = 0 \quad (\forall \nu \neq \mu \wedge \forall \nu' \neq \mu).$$

Proof of Lemma A.4. The proof is straightforward by using Lemmas A.2, A.3 and equations (A.9), (A.10). □

To derive the elements of the product of matrices $\Omega_i \Omega^{-1} \Omega_j$, needed for the partial derivatives of matrix A (see Lemmas A.14 – A.17), we define the following matrices:

$$\begin{aligned} W_{ij} \;=\; & \sigma^{i\mu}\sigma_{\mu\mu'}\sigma^{\mu' j}R^{i\mu}_{\rho_\mu}R_{\mu\mu'}R^{\mu' j}_{\rho_{\mu'}} \\ & + \delta_{\mu i}\left\{\left[\sum_{k=1}^{M}\sigma^{\mu k}\sigma_{k\mu'}R^{\mu k}_{\rho_\mu}R_{k\mu'}\right] + \sigma^{\mu\mu}\sigma_{\mu\mu'}R^{\mu\mu}_{\rho_\mu}R_{\mu\mu'}\right\}\sigma^{\mu' j}R^{\mu' j}_{\rho_{\mu'}} \\ & + \delta_{j\mu'}\sigma^{i\mu}R^{i\mu}_{\rho_\mu}\left\{\left[\sum_{r=1}^{M}\sigma_{\mu r}\sigma^{r\mu'}R_{\mu r}R^{r\mu'}_{\rho_{\mu'}}\right] + \sigma_{\mu\mu'}\sigma^{\mu'\mu'}R_{\mu\mu'}R^{\mu'\mu'}_{\rho_{\mu'}}\right\} \\ & + \delta_{\mu i}\delta_{j\mu'}\left\{\sum_{k=1}^{M}\sum_{r=1}^{M}\sigma^{\mu k}\sigma_{kr}\sigma^{r\mu'}R^{\mu k}_{\rho_\mu}R_{kr}R^{r\mu'}_{\rho_{\mu'}}\right. \\ & + \left[\sum_{k=1}^{M}\sigma^{\mu k}\sigma_{k\mu'}R^{\mu k}_{\rho_\mu}R_{k\mu'}\right]\sigma^{\mu'\mu'}R^{\mu'\mu'}_{\rho_{\mu'}} \\ & + \sigma^{\mu\mu}R^{\mu\mu}_{\rho_\mu}\left[\sum_{r=1}^{M}\sigma_{\mu r}\sigma^{r\mu'}R_{\mu r}R^{r\mu'}_{\rho_{\mu'}}\right] \\ & \left. + \sigma^{\mu\mu}\sigma_{\mu\mu'}\sigma^{\mu'\mu'}R^{\mu\mu}_{\rho_\mu}R_{\mu\mu'}R^{\mu'\mu'}_{\rho_{\mu'}}\right\}, \end{aligned} \tag{A.14}$$

$$\Omega^*_{\rho_\mu \rho_{\mu'}} = \Omega_{\rho_\mu}\Omega^{-1}\Omega_{\rho_{\mu'}}, \quad \Omega^*_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}} = \Omega_{\varsigma_{(\mu\mu')}}\Omega^{-1}\Omega_{\varsigma_{(\nu\nu')}}, \tag{A.15}$$

$$\Omega^*_{\rho_\mu \varsigma_{(\nu\nu')}} = \Omega_{\rho_\mu}\Omega^{-1}\Omega_{\varsigma_{(\nu\nu')}} \quad \text{and} \quad \Omega^*_{\varsigma_{(\nu\nu')}\rho_\mu} = \Omega_{\varsigma_{(\nu\nu')}}\Omega^{-1}\Omega_{\rho_\mu}.$$

Lemma A.5. *The elements of matrices* $\Omega^*_{\rho_\mu \rho_{\mu'}}$, $\Omega^*_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}}$, $\Omega^*_{\rho_\mu \varsigma_{(\nu\nu')}}$ *and* $\Omega^*_{\varsigma_{(\nu\nu')}\rho_\mu}$ *can be analytically written as follows:*

$$\begin{aligned} \Omega^*_{\rho_\mu \rho_{\mu'}} &= [(W_{ij})_{i,j=1,\,\dots,\,M}], \\ \Omega^*_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}} &= [(\delta_{\mu i}\delta_{j\nu'}\sigma_{\mu'\nu}R^{\mu\nu'})_{i,j=1,\,\dots,\,M}], \\ \Omega^*_{\rho_\mu \varsigma_{(\nu\nu')}} &= \left[\left(\left(\sum_{k=1}^{M}\sigma^{ik}\sigma_{k\nu}R^{ik}_{\rho_\mu}R_{k\nu}\right)\delta_{j\nu'}R^{\nu\nu'}\right)_{i,j=1,\,\dots,\,M}\right], \\ \Omega^*_{\varsigma_{(\nu\nu')}\rho_\mu} &= \left[\left(\delta_{\nu i}R^{\nu\nu'}\left(\sum_{r=1}^{M}\sigma_{\nu' r}\sigma^{rj}R_{\nu' r}R^{rj}_{\rho_\mu}\right)\right)_{i,j=1,\,\dots,\,M}\right]. \end{aligned} \tag{A.16}$$

Proof of Lemma A.5. The results (A.16) can be easily proved by using Lemma A.4 and equations (A.9), (A.14). □

Asymptotic expansions of estimators

For all estimators of matrix B and the nuisance patameters considered in the paper, in the next lemmas we derive the following asymptotic expansions. In each case, these estimators are indexed by I (see footnotes 2 and 3).

Lemma A.6. *All estimators* $\hat{B}_I$ $(I = UL, RL, GL, IG, ML)$ *of matrix* B*, defined in (21), admit a stochastic expansion of the form*

$$\hat{B}_I = B + \tau B_1^I + \omega(\tau^2), \tag{A.17}$$

where

$$\begin{aligned} B_1^{UL} &= \sqrt{T}(Z'Z)^{-1}Z'E, \\ vec(B_1^{RL}) &= \sqrt{T}\Psi(X_*'X_*)^{-1}X_*'\varepsilon, \\ vec(B_1^{GL}) &= vec(B_1^{IG}) = vec(B_1^{ML}) \\ &= \sqrt{T}\Psi\left[X_*'(\Sigma_I^{-1}\otimes I_T)X_*\right]^{-1}X_*'(\Sigma_I^{-1}\otimes I_T)\varepsilon. \end{aligned} \tag{A.18}$$

Proof of Lemma A.6. The results of the lemma follow immediately by using the definitions of all estimators B_I considered (see footnote 3). □

Let $\hat{E}_I$ be the residuals corresponding to the estimators $\hat{B}_I$. Then, the following lemma holds for the estimators $\hat{\Sigma}_I$ and $\hat{\Sigma}_I^{-1}$ of matrix Σ and its inverse, respectively, based on $\hat{E}_I$.

Lemma A.7. *All estimators* $\hat{\Sigma}_I$ $(I = UL, RL, GL, IG, ML)$ *of matrix* Σ *admit a stochastic expansion of the form*

$$\hat{\Sigma}_I = \Sigma + \tau(\Sigma_1 + \tau\Sigma_2^I) + \omega(\tau^3), \tag{A.19}$$

where

$$\Sigma_1 = \sqrt{T}(E'E/T - \Sigma), \quad \Sigma_2^I = (B_1^I - B_1^{UL})'\Gamma(B_1^I - B_1^{UL}) - E'P_Z E, \tag{A.20}$$

Γ is any conformable matrix and P_Z is the orthogonal projector spanned by the columns of matrix Z. Estimator $\hat{\Sigma}_I^{-1}$ admits a stochastic expansion of the form

$$\hat{\Sigma}_I^{-1} = \Sigma^{-1} - \tau S_1 + \tau^2 S_2^I + \omega(\tau^3), \tag{A.21}$$

where

$$S_1 = \Sigma^{-1}\Sigma_1\Sigma^{-1}, \quad S_2^I = \Sigma^{-1}(\Sigma_1\Sigma^{-1}\Sigma_1 - \Sigma_2^I)\Sigma^{-1}. \tag{A.22}$$

Proof of Lemma A.7. The proof is straightforward based on Lemma A.6. □

The stochastic expansion of estimator of vector ς, denoted as $\hat{\varsigma}_I$ is given in the next lemma.

Lemma A.8. *All estimators $\hat{\varsigma}_I = vec\{[\hat{E}_I'\hat{E}_I/T]^{-1}\}$ ($I = UL$, RL, GL, IG, ML) of vector ς admit a stochastic expansion of the form*

$$\hat{\varsigma}_I = \varsigma - \tau vec(S_1) + \tau^2 vec(S_2^I) + \omega(\tau^3) \tag{A.23}$$

and thus, the $(M^2 \times 1)$ vector $\delta_\varsigma = (\hat{\varsigma} - \varsigma)/\tau$, with elements $\delta_{\varsigma_{(\mu\mu')}}$ defined in (27), admits a stochastic expansion of the form

$$\begin{aligned}\delta_\varsigma &= -vec(S_1) + \tau vec(S_2^I) + \omega(\tau^2) \\ &= d_{1\varsigma} + \tau d_{2\varsigma} + \omega(\tau^2),\end{aligned} \tag{A.24}$$

which implies that

$$d_{1\varsigma} = -vec(S_1), \quad d_{2\varsigma} = vec(S_2^I). \tag{A.25}$$

Proof of Lemma A.8. The proof follows simply from equations (21), (29), (32) and (A.21). □

To derive the stochastic expansion of the estimators of σ, denoted as $\hat{\sigma}_I$, we define the following $(M \times M)$ matrices (indexed by I):

$$\begin{aligned}\Delta_I &= \lim_{T\to\infty} T\mathcal{E}[(\hat{B}_I - \hat{B}_{UL})'\Gamma(\hat{B}_I - \hat{B}_{UL})] \\ &= \lim_{T\to\infty} \mathcal{E}[(B_1^I - B_1^{UL})'\Gamma(B_1^I - B_1^{UL})], \qquad (A.26)\end{aligned}$$

where Γ is any conformable matrix.

Lemma A.9. *All estimators* $\hat{\sigma}_I^2$ $(I = UL, RL, GL, IG, ML)$ *of* σ^2 *(see footnote 1) satisfy the relation*

$$\begin{aligned}\hat{\sigma}_I^2 &= tr(\hat{\Sigma}_I^{-1}\hat{\Sigma}_J)/(M - \tau^2 n) \\ &= \{M + \tau^2 tr[(S_2^I - S_2^J)\Sigma]\}/(M - \tau^2 n) + \omega(\tau^3). \qquad (A.27)\end{aligned}$$

The last equation implies that

$$\begin{aligned}(\hat{\sigma}_I^2 - 1)/\tau &= \{M/\tau + \tau tr[(S_2^I - S_2^J)\Sigma]\}/(M - \tau^2 n) - 1/\tau + \omega(\tau^2) \\ &= \tau\{tr[(S_2^I - S_2^J)\Sigma] + n\}/M + \omega(\tau^2), \qquad (A.28)\end{aligned}$$

i.e., scalar δ_0*, defined in (27), admits a stochastic expansion of the form*

$$\delta_0 = \sigma_0 + \tau\sigma_1 + \omega(\tau^2), \qquad (A.29)$$

which in turn implies that

$$\sigma_0 = 0 \quad and \quad \sigma_1 = \{tr[(S_2^I - S_2^J)\Sigma] + n\}/M. \qquad (A.30)$$

Proof of Lemma A.9. To prove the lemma we rely on the following results (see (A.31) and (A.32)): Since the rows ε_t $(t = 1, \ldots, T)$ of E are independent $\mathcal{N}_M(0, \Sigma)$ random vectors, matrix $E'E$ is a Wishart matrix with weight matrix Σ and T degrees of freedom, i.e., $E'E \sim \mathcal{W}_M(\Sigma, T)$ and $\mathcal{E}(E'E) = T\Sigma$. Then, it is easy to show that

$$\mathcal{E}(E'E\Sigma^{-1}E'E) = T(M + T + 1)\Sigma. \qquad (A.31)$$

Moreover, since $E'E \sim \mathcal{W}_M(\Sigma, T)$ and P_Z is idempotent of rank K, it follows that matrix $E'P_ZE \sim \mathcal{W}_M(\Sigma, K)$ and $\mathcal{E}(E'P_ZE) = tr(P_Z)\Sigma = K\Sigma$. Further, $\mathcal{E}(\Sigma_1) = 0$,

$\mathcal{E}(\Sigma_1\Sigma^{-1}\Sigma_1) = (M+1)\Sigma$ and

$$\mathcal{E}(S_1) = 0,\ \mathcal{E}(S_2^I) = (M+K+1)\Sigma^{-1} - \Sigma^{-1}\mathcal{E}[(B_1^I - B_1^{UL})'\Gamma(B_1^I - B_1^{UL})]\Sigma^{-1}. \quad \text{(A.32)}$$

Let $\hat{\varepsilon}_{GL} = vec(\hat{E}_{GL})$ be the GL residuals of regression equation (16). Then, the corresponding estimator of Σ is $\hat{\Sigma}_J = \hat{E}'_{GL}\hat{E}_{GL}/T$. Also, let $\hat{\beta}_{GL}$ be the GL estimator of β in (16). Define the $(M \times M)$ matrices $M_I = \lim_{T\to\infty} \mathcal{E}(S_2^I)$ ($I = UL$, RL, GL, IG, ML) and the $(M^2 \times M^2)$ matrix N whose $((ij),(kr))$-th element is $\nu_{(ij)(kr)} = \sigma_{ik}\sigma_{jr} + \sigma_{ir}\sigma_{jk}$ ($i,j,k,r = 1,\ \ldots,\ M$). Then, (A.26) and (A.32) imply that

$$M_I = (M+K+1)\Sigma^{-1} - \Sigma^{-1}\Delta_I\Sigma^{-1} \quad \text{(A.33)}$$

$$\Rightarrow \lim_{\to\infty} T\mathcal{E}[(S_2^I - S_2^J)\Sigma] = (M_I - M_{GL})\Sigma = \Sigma^{-1}(\Delta_{GL} - \Delta_I), \quad \text{(A.34)}$$

where

$$\Delta_{UL} = 0,$$

$$\Delta_{RL} = \left[\left[\left(tr(B_{ii}^{-1}B_{ij}B_{jj}^{-1}B_{ji}) - n_i - n_j + K\right)\sigma_{ij}\right]_{i,j=1,\ \ldots,\ M}\right], \quad \text{(A.35)}$$

$$\Delta_{GL} = \Delta_{IG} = \Delta_{ML} = K\Sigma - \left[(tr(G_{ij}B_{ji}))_{i,j=1,\ \ldots,\ M}\right].$$

Since $E'E \sim \mathcal{W}_M(\Sigma,T)$ and $\mathcal{E}(E'E) = T\Sigma$, matrix $W_* = \sqrt{T}\Sigma_1 = E'E - T\Sigma$, with elements w_{ij}, is a Wishart matrix in deviations from it expected values. Following Zellner (1971, page 389, equation (B.58)), we find that

$$\mathcal{E}(w_{ij}w_{kr}) = T(\sigma_{ik}\sigma_{jr} + \sigma_{ir}\sigma_{jk}) = T\nu_{(ij)(kr)} \quad \text{(A.36)}$$

$$\Rightarrow \lim_{T\to\infty} \mathcal{E}[(vec(S_1))(vec(S_1))'] = (\Sigma^{-1}\otimes\Sigma^{-1})N(\Sigma^{-1}\otimes\Sigma^{-1}). \quad \text{(A.37)}$$

The proof of the lemma can be completed using the following relationship:

$$(M - \tau^2 n)^{-1} = M^{-1}(1 - \tau^2 n/M)^{-1} = (1 + \tau^2 n/M)/M + \omega(\tau^4). \quad \text{(A.38)}$$

□

Before deriving the asymptotic expansion of the estimators of ρ_μ, next we define the following $(T \times T)$ matrices:

$$R_i^{\mu\mu} = R_{\rho_\mu}^{\mu\mu} + i\rho_\mu \Delta \quad (i = 1, 2),$$

$$V_\mu = \left[I - X_\mu (X_\mu' R^{\mu\mu} X_\mu)^{-1} X_\mu' R_{\mu\mu}\right] R^{\mu\mu}. \tag{A.39}$$

The first assumption in Subsection 3.1 implies that matrices $X_\mu' R^{\mu\mu} X_\mu / T$ and $X_\mu' X_\mu / T$ converge to non-singular matrices, as $T \to \infty$, and that matrices

$$X_\mu' \Delta X_\mu / T, \quad X_\mu' \Delta R_{\mu\mu} X_\mu / T, \quad X_\mu' R_{\mu\mu} \Delta X_\mu / T, \quad X_\mu' \Delta R_{\mu\mu} \Delta X_\mu / T \quad \text{and} \quad X_\mu' R_{\mu\mu} X_\mu / T$$

are of order $O(T^{-1})$. The above matrices help to derive expectations of products of quadratic forms of u, needed in the expansions of estimators of ρ_μ. These are given in the next lemma:

Lemma A.10. *For quadratic forms of vector u, we have the following results:*

$$\begin{aligned}
\mathcal{E}(u_\mu' R_2^{\mu\mu} u_\mu) &= \frac{2\rho_\mu \sigma_{\mu\mu}}{1 - \rho_\mu^2}, \\
\mathcal{E}(u_\mu' u_\mu u_\mu' R_2^{\mu\mu} u_\mu) &= -\frac{2T\rho_\mu \sigma_{\mu\mu}^2}{(1 - \rho_\mu^2)^2} + O(1), \\
\mathcal{E}(u_\mu' R_2^{\mu\mu} u_\mu u_\mu' R_2^{\mu\mu} u_\mu) &= \frac{4T\sigma_{\mu\mu}^2}{1 - \rho_\mu^2} + O(1), \\
\mathcal{E}(u_\mu' R_2^{\mu\mu} u_\mu u_{\mu'}' R_2^{\mu'\mu'} u_{\mu'}) &= \frac{4T\sigma_{\mu\mu}\sigma_{\mu'\mu'}}{1 - \rho_\mu \rho_{\mu'}} + O(1), \\
\mathcal{E}(u_\mu' \overline{P}_{X_\mu} R_2^{\mu\mu} P_{X_\mu} V_\mu R^{\mu\mu} u_\mu) &= \frac{\sigma_{\mu\mu}}{\rho_\mu}\left[n_\mu - tr(F_{\mu\mu}^{-1} B_{\mu\mu} F_{\mu\mu}^{-1} \Theta_{\mu\mu})\right] \\
&\quad + O(T^{-1}), \qquad\qquad (A.40) \\
\mathcal{E}(u_\mu' \overline{P}_{X_\mu} R_2^{\mu\mu} \overline{P}_{X_\mu} u_\mu) &= \frac{\sigma_{\mu\mu}}{\rho_\mu}\Big[2\left[\rho_\mu^2/(1 - \rho_\mu^2) - n_\mu\right] \\
&\quad + (1 - \rho_\mu^2) tr(F_{\mu\mu}^{-1}\Theta_{\mu\mu}) \\
&\quad + tr(F_{\mu\mu}^{-1} B_{\mu\mu} F_{\mu\mu}^{-1} \Theta_{\mu\mu})\Big] + O(T^{-1}), \\
\mathcal{E}(u_\mu' R^{\mu\mu} V_\mu P_{X_\mu} R_2^{\mu\mu} P_{X_\mu} V_\mu R^{\mu\mu} u_\mu) &= \frac{\sigma_{\mu\mu}}{\rho_\mu}\Big[\left[tr(F_{\mu\mu}^{-1} B_{\mu\mu} F_{\mu\mu}^{-1} \Theta_{\mu\mu}) - n_\mu\right] \\
&\quad + (1 - \rho_\mu^2)\left[tr(F_{\mu\mu} B_{\mu\mu}^{-1}) - tr(F_{\mu\mu}^{-1}\Theta_{\mu\mu})\right]\Big] \\
&\quad + O(T^{-1}).
\end{aligned}$$

Proof of Lemma A.10. The results of the lemma follow based on the result of Magnus and Neudecker (1979) given in page 389, after tedious algebra. Note that in calculating the traces of the lemma, terms of the form $T^n \rho_\mu^{2T} \to 0$ since $0 \leq \rho_\mu < 1$ and L' Hospital's rule implies that $\lim_{T\to\infty} T^n \rho_\mu^{2T} = 0$. □

The stochastic expansion of the LS estimator of ρ_μ is given in the next lemma:

Lemma A.11. *The LS estimator of ρ_μ, denoted as $\tilde{\rho}_\mu$, admits a stochastic expansion of the form*

$$\tilde{\rho}_\mu = \rho_\mu + \tau\left(\rho_\mu^{(1)} + \tau\rho_\mu^{(2)}\right) + \omega(\tau^3), \tag{A.41}$$

where

$$\rho_\mu^{(1)} = -\frac{u'_\mu R_2^{\mu\mu} u_\mu}{2\sqrt{T}\sigma_{u_\mu}^2}, \quad \rho_\mu^{(2)} = -\frac{u'_\mu \overline{P}_{X_\mu} R_2^{\mu\mu} \overline{P}_{X_\mu} u_\mu}{2\sigma_{u_\mu}^2} + \frac{u'_\mu u_\mu u'_\mu R_2^{\mu\mu} u_\mu}{2T\sigma_{u_\mu}^4}. \tag{A.42}$$

Proof of Lemma A.11. To prove the lemma, we rely on the following results (see (A.43) – (A.46)): Let ε_{ti} be the (t,i)-th element of matrix E. Then, the (i,j)-th element of matrix $E'E/T$ is

$$e_{ij} = \sum\nolimits_{t=1}^{T} \varepsilon_{ti}\varepsilon_{tj}/T = \varepsilon'_i\varepsilon_j/T, \tag{A.43}$$

where ε_i is the i-th column of matrix E. Since σ_{ij} and σ^{ij} are the (i,j)-th elements of matrices Σ and Σ^{-1}, respectively, $\Sigma^{-1} = \Sigma^{-1}\Sigma\Sigma^{-1}$ implies that $\sigma^{ij} = \sum_{k=1}^{M}\sum_{r=1}^{M}\sigma^{ik}\sigma_{kr}\sigma^{rj}$. Thus, the (i,j)-th element of matrix Σ_1 in Lemma A.7 is given as

$$\sigma_{ij}^{(1)} = \sqrt{T}\left(e_{ij} - \sigma_{ij}\right) \tag{A.44}$$

and the (ij)-th element of $(M^2 \times 1)$ vector $vec(S_1)$, where $S_1 = \Sigma^{-1}\Sigma_1\Sigma^{-1}$, is given as

$$s_{(ij)}^{(1)} = \sqrt{T}\left\{\sum\nolimits_{k=1}^{M}\sum\nolimits_{r=1}^{M}\sigma^{ik}(\varepsilon'_k\varepsilon_r/T)\sigma^{rj} - \sigma^{ij}\right\}. \tag{A.45}$$

Since $u_\mu = P_\mu\varepsilon_\mu \Rightarrow u'_\mu R_2^{\mu\mu} u_\mu = \varepsilon'_\mu P'_\mu R_2^{\mu\mu} P_\mu \varepsilon_\mu$ and $R_{\mu\mu} = P_\mu P'_\mu$, we can show that

$$\mathcal{E}(u'_\mu R_2^{\mu\mu} u_\mu) = \sigma_{\mu\mu} tr(R_2^{\mu\mu} R_{\mu\mu}) \Rightarrow$$

$$\Rightarrow \mathcal{E}[(\varepsilon_k'\varepsilon_r/T)u_\mu' R_2^{\mu\mu} u_\mu] = \sigma_{kr}\sigma_{\mu\mu}\frac{2\rho_\mu}{1-\rho_\mu^2} + O(T^{-1})$$

$$\Rightarrow \mathcal{E}\left(s_{(ij)}^{(1)} u_\mu' R_2^{\mu\mu} u_\mu\right) = \sqrt{T}\sigma_{\mu\mu}\frac{2\rho_\mu}{1-\rho_\mu^2}\left\{\sum\nolimits_{k=1}^{M}\sum\nolimits_{r=1}^{M}\sigma^{ik}\sigma_{kr}\sigma^{rj} - \sigma^{ij}\right\} + O(T^{-1/2})$$

$$\Rightarrow \lim_{T\to\infty}\mathcal{E}\left(s_{(ij)}^{(1)} u_\mu' R_2^{\mu\mu} u_\mu\right) = 0. \tag{A.46}$$

The rest of the proof follows using Lemma A.10. □

The stochastic expansions of the rest of the estimators of ρ_μ, listed in footnote 2, are given in the next lemma.

Lemma A.12. *The GL, PW, ML and DW estimators of ρ_μ admit the following stochastic expansions, respectively:*

$$\begin{aligned}
\hat{\rho}_\mu^{GL} = \hat{\rho}_\mu^{PW} &= \tilde{\rho}_\mu - \tau^2\frac{1-\rho_\mu^2}{\sigma_{\mu\mu}}\left[u_\mu'\overline{P}_{X_\mu}R_2^{\mu\mu}P_{X_\mu}V_\mu R^{\mu\mu}u_\mu\right.\\
&\quad \left.+u_\mu' R^{\mu\mu}V_\mu P_{X_\mu}R_2^{\mu\mu}P_{X_\mu}V_\mu R^{\mu\mu}u_\mu/2\right] + \omega(\tau^3),\\
\hat{\rho}_\mu^{ML} &= \hat{\rho}_\mu^{GL} + \tau^2\left[\rho_\mu\frac{1-\rho_\mu^2}{\sigma_{\mu\mu}}(u_{1\mu}^2 + u_{T\mu}^2) - \rho_\mu\right] + \omega(\tau^3),\\
\hat{\rho}_\mu^{DW} &= \tilde{\rho}_\mu + \tau^2\frac{1-\rho_\mu^2}{2\sigma_{\mu\mu}}(u_{1\mu}^2 + u_{T\mu}^2) + \omega(\tau^3).
\end{aligned} \tag{A.47}$$

Proof of Lemma A.12. The results of the lemma can be easily proved based on Magee (1985, pages 279–281) for the GL and iterative PW estimators of ρ_μ, Beach and MacKinnon (1978, pages 52–54) and Magee (1985, pages 281–284) for the ML estimator, and using Lemma A.11 and the definition of the DW estimator of ρ_μ. □

The stochastic expansion of the elements of vector δ_ϱ, are given in the next lemma:

Lemma A.13. *The $(M \times 1)$ vector $\delta_\varrho = \sqrt{T}(\hat{\varrho} - \varrho)/\tau$, with elements δ_{ρ_μ} defined in (27), admits a stochastic expansion of the form*

$$\delta_\varrho = d_{1\varrho} + \tau d_{2\varrho} + \omega(\tau^2). \tag{A.48}$$

For estimators $\hat{\rho}_\mu^I$ $(I = LS, GL, PW, ML, DW)$*, the elements of* $d_{1\varrho}$ *and* $d_{1\varrho}$ *in* (A.48) *are analytically given as follows:* $d^{GL}_{(1)\rho_\mu} = d^{PW}_{(1)\rho_\mu} = d^{ML}_{(1)\rho_\mu} = d^{DW}_{(1)\rho_\mu} = d^{LS}_{(1)\rho_\mu}$ *and*

$$
\begin{aligned}
d^{LS}_{(1)\rho_\mu} &= \rho_\mu^{(1)}, \\
d^{LS}_{(2)\rho_\mu} &= \rho_\mu^{(2)}, \\
d^{GL}_{(2)\rho_\mu} = d^{PW}_{(2)\rho_\mu} &= d^{LS}_{(2)\rho_\mu} - \frac{1-\rho_\mu^2}{\sigma_{\mu\mu}} \left[u'_\mu \overline{P}_{X_\mu} R_2^{\mu\mu} P_{X_\mu} V_\mu R^{\mu\mu} u_\mu \right. \\
&\quad \left. + u'_\mu R^{\mu\mu} V_\mu P_{X_\mu} R_2^{\mu\mu} P_{X_\mu} V_\mu R^{\mu\mu} u_\mu / 2 \right], \qquad \text{(A.49)} \\
d^{ML}_{(2)\rho_\mu} &= d^{GL}_{(2)\rho_\mu} + \rho_\mu \frac{1-\rho_\mu^2}{\sigma_{\mu\mu}} (u_{1\mu}^2 + u_{T\mu}^2) - \rho_\mu, \\
d^{DW}_{(2)\rho_\mu} &= d^{LS}_{(2)\rho_\mu} + \frac{1-\rho_\mu^2}{2\sigma_{\mu\mu}} (u_{1\mu}^2 + u_{T\mu}^2).
\end{aligned}
$$

Proof of Lemma A.13. The proof is straightforward using Lemmas A.11 and A.12.

□

Next, we provide analytic forms of the elements of vectors l and c, and matrices L, C and D_* employed in the stochastic expansions of the tests statistics given in the paper. To this end, we first derive the partial derivatives of matrix A, given in (28), with respect to the elements of ϱ and ς. Using matrices $B_{ij} = X'_i R^{ij} X_j / T$, matrix A defined in (28) can be partitioned as follows:

$$A = [(\sigma^{ij} B_{ij})_{i,j=1,\, \dots,\, M}]. \qquad \text{(A.50)}$$

Lemma A.14. *The partial derivatives of matrix* A *with respect to the elements of* ϱ

and ς can be analytically written as follows:

$$A_{\rho_\mu} = [(\frac{\sigma^{ij}}{T} X_i' R_{\rho_\mu}^{ij} X_j)_{i,j=1,\dots,M}], \quad A_{\rho_\mu \rho_{\mu'}} = [(\frac{\sigma^{ij}}{T} X_i' R_{\rho_\mu \rho_{\mu'}}^{ij} X_j)_{i,j=1,\dots,M}],$$

$$A_{\rho_\mu \rho_{\mu'}}^* = [(X_i' W_{ij} X_j / T)_{i,j=1,\dots,M}], \quad A_{\varsigma_{(\mu\mu')}} = [(\delta_{\mu i} \delta_{j\mu'} B_{\mu\mu'})_{i,j=1,\dots,M}],$$

$$A_{\varsigma_{(\mu\mu')} \varsigma_{(\nu\nu')}} = 0, \quad A_{\varsigma_{(\mu\mu')} \varsigma_{(\nu\nu')}}^* = \sigma_{\mu'\nu} A_{\varsigma_{(\mu\nu')}},$$

$$A_{\rho_\mu \varsigma_{(\nu\nu')}} = [(\delta_{\nu i} \delta_{j\nu'} X_\nu' R_{\rho_\mu}^{\nu\nu'} X_{\nu'} / T)_{i,j=1,\dots,M}], \tag{A.51}$$

$$A_{\rho_\mu \varsigma_{(\nu\nu')}}^* = \left[\left(\sum_{k=1}^M \frac{\delta_{j\nu'} \sigma^{ik} \sigma_{k\nu}}{T} X_i' R_{\rho_\mu}^{ik} R_{k\nu} R^{\nu\nu'} X_{\nu'} \right)_{i,j=1,\dots,M} \right],$$

$$A_{\varsigma_{(\nu\nu')} \rho_\mu}^* = \left[\left(\sum_{r=1}^M \frac{\delta_{\nu i} \sigma_{\nu' r} \sigma^{rj}}{T} X_\nu' R^{\nu\nu'} R_{\nu' r} R_{\rho_\mu}^{rj} X_j \right)_{i,j=1,\dots,M} \right].$$

Proof of Lemma A.14. The proof follows immediately from equation (31), and Lemmas A.4 and A.5. □

Analytic formulae of the elements of vector l and matrix L are given in the following lemma

Lemma A.15. *The elements of vector l and matrix L can be calculated as follows:*

$$l_{\rho_\mu} = \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M \sum_{r=1}^M \sigma^{kr} h_i' G_{ik} X_k' R_{\rho_\mu}^{kr} X_r G_{rj} h_j / T, \tag{A.52}$$

$$l_{\varsigma_{(\mu\mu')}} = \sum_{i=1}^M \sum_{j=1}^M h_i' G_{i\mu} B_{\mu\mu'} G_{\mu' j} h_j, \tag{A.53}$$

$$\begin{aligned} l_{\rho_\mu \rho_{\mu'}} \quad = \quad & \sum_{q=1}^M \sum_{s=1}^M \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M \sum_{r=1}^M \sigma^{ik} \sigma^{rj} \\ & \times h_q' G_{qi} X_i' R_{\rho_\mu}^{ik} (\sigma_{kr} R_{kr} - 2 X_k G_{kr} X_r' / T) R_{\rho_{\mu'}}^{rj} X_j G_{js} h_s / T \\ & + \sum_{q=1}^M \sum_{s=1}^M \sum_{i=1}^M \sum_{j=1}^M \sigma^{ij} \\ & \times h_q' G_{qi} X_i' R_{\rho_\mu \rho_{\mu'}}^{ij} X_j G_{js} h_s / 2T, \end{aligned} \tag{A.54}$$

$$l_{\varsigma_{(\mu\mu')} \varsigma_{(\nu\nu')}} = \sigma_{\mu'\nu} l_{\varsigma_{(\mu\nu')}} - 2 \sum_{i=1}^M \sum_{j=1}^M h_i' G_{i\mu} B_{\mu\mu'} G_{\mu'\nu} B_{\nu\nu'} G_{\nu' j} h_j, \tag{A.55}$$

$$\begin{aligned} l_{\rho_\mu \varsigma_{(\nu\nu')}} \quad = \quad & \sum_{q=1}^M \sum_{s=1}^M \sum_{i=1}^M \sum_{k=1}^M \sigma^{ik} h_q' G_{qi} X_i' R_{\rho_\mu}^{ik} \\ & \times (\sigma_{k\nu} R_{k\nu} - 2 X_k G_{k\nu} X_\nu' / T) R^{\nu\nu'} X_{\nu'} G_{\nu' s} h_s / T \\ & + \sum_{q=1}^M \sum_{s=1}^M h_q' G_{q\nu} X_\nu' R_{\rho_\mu}^{\nu\nu'} X_{\nu'} G_{\nu' s} h_s / 2T, \end{aligned} \tag{A.56}$$

$$l_{\varsigma_{(\nu\nu')}\rho_\mu} = \sum_{q=1}^{M}\sum_{s=1}^{M}\sum_{j=1}^{M}\sum_{r=1}^{M} \sigma^{rj} h_q' G_{q\nu} X_\nu' R^{\nu\nu'} \times(\sigma_{\nu' r} R_{\nu' r} - 2X_{\nu'} G_{\nu' r} X_r'/T) R_{\rho_\mu}^{rj} X_j G_{js} h_s/T + \sum_{q=1}^{M}\sum_{s=1}^{M} h_q' G_{q\nu} X_\nu' R_{\rho_\mu}^{\nu\nu'} X_{\nu'} G_{\nu' s} h_s/2T. \quad \text{(A.57)}$$

Proof of Lemma A.15. The results of the lemma follow by using the definitions in (41), partition of the matrix G in (52) and Lemmas A.1 – A.14. □

Analytic formulae of the elements of vector c and matrices C and D_* are given in the following lemma:

Lemma A.16. *The elements of vector c and matrices C and D_* can be calculated as follows:*

$$c_{\rho_\mu} = \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{ij} tr(X_i' R_{\rho_\mu}^{ij} X_j \Xi_{ji})/T, \quad \text{(A.58)}$$

$$c_{\varsigma_{(\mu\mu')}} = tr(B_{\mu\mu'} \Xi_{\mu'\mu}), \quad \text{(A.59)}$$

$$c_{\rho_\mu\rho_{\mu'}} = \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{r=1}^{M} \sigma^{ik}\sigma_{kr}\sigma^{rj} \times tr(X_i' R_{\rho_\mu}^{ik} R_{kr} R_{\rho_{\mu'}}^{rj} X_j \Xi_{ji})/T - 2\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{r=1}^{M} \sigma^{ik}\sigma^{rj} \times tr(X_i' R_{\rho_\mu}^{ik} X_k G_{kr} X_r' R_{\rho_{\mu'}}^{rj} X_j \Xi_{ji})/T^2 + \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{ij} tr(X_i' R_{\rho_\mu\rho_{\mu'}}^{ij} X_j \Xi_{ji})/T, \quad \text{(A.60)}$$

$$c_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}} = \sigma_{\mu'\nu} c_{\varsigma_{(\mu\nu')}} - 2tr(B_{\mu\mu'} G_{\mu'\nu} B_{\nu\nu'} \Xi_{\nu'\mu}), \quad \text{(A.61)}$$

$$c_{\rho_\mu\varsigma_{(\nu\nu')}} = \sum_{i=1}^{M}\sum_{k=1}^{M} \sigma^{ik}\sigma_{k\nu} tr(X_i' R_{\rho_\mu}^{ik} R_{k\nu} R^{\nu\nu'} X_{\nu'} \Xi_{\nu' i})/T - 2\sum_{i=1}^{M}\sum_{k=1}^{M} \sigma^{ik} tr(X_i' R_{\rho_\mu}^{ik} X_k G_{k\nu} B_{\nu\nu'} \Xi_{\nu' i})/T + tr(X_\nu' R_{\rho_\mu}^{\nu\nu'} X_{\nu'} \Xi_{\nu'\nu})/2T, \quad \text{(A.62)}$$

$$c_{\varsigma_{(\nu\nu')}\rho_\mu} = \sum_{j=1}^{M}\sum_{r=1}^{M} \sigma_{\nu' r}\sigma^{rj} tr(X_\nu' R^{\nu\nu'} R_{\nu' r} R_{\rho_\mu}^{rj} X_j \Xi_{j\nu})/T - 2\sum_{j=1}^{M}\sum_{r=1}^{M} \sigma^{rj} tr(B_{\nu\nu'} G_{\nu' r} X_r' R_{\rho_\mu}^{rj} X_j \Xi_{j\nu})/T + tr(X_\nu' R_{\rho_\mu}^{\nu\nu'} X_{\nu'} \Xi_{\nu'\nu})/2T, \quad \text{(A.63)}$$

$$d_{\rho_\mu \rho_{\mu'}} = \sum\nolimits_{i=1}^{M} \sum\nolimits_{j=1}^{M} \sum\nolimits_{k=1}^{M} \sum\nolimits_{r=1}^{M} \sigma^{ik} \sigma^{rj} \times tr(X_i' R_{\rho_\mu}^{ik} X_k \Xi_{kr} X_r' R_{\rho_{\mu'}}^{rj} X_j \Xi_{ji})/2T^2, \quad (A.64)$$

$$d_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}} = tr(B_{\mu\mu'} \Xi_{\mu'\nu} B_{\nu\nu'} \Xi_{\nu'\mu})/2, \quad (A.65)$$

$$d_{\rho_\mu \varsigma_{(\nu\nu')}} = \sum\nolimits_{i=1}^{M} \sum\nolimits_{k=1}^{M} \sigma^{ik} tr(X_i' R_{\rho_\mu}^{ik} X_k \Xi_{k\nu} B_{\nu\nu'} \Xi_{\nu' i})/2T, \quad (A.66)$$

$$d_{\varsigma_{(\nu\nu')}\rho_\mu} = \sum\nolimits_{j=1}^{M} \sum\nolimits_{r=1}^{M} \sigma^{rj} tr(B_{\nu\nu'} \Xi_{\nu' r} X_r' R_{\rho_\mu}^{rj} X_j \Xi_{j\nu})/2T. \quad (A.67)$$

Proof of Lemma A.16. The results of the lemma can be easily calculated by using the definitions (56) and (57), partition of matrix Ξ in (52) and the following traces:

$$\begin{gathered} tr(A_{\rho_\mu}\Xi), \quad tr(A_{\rho_\mu\rho_{\mu'}}\Xi), \quad tr(A_{\varsigma_{(\mu\mu')}}\Xi), \quad tr(A_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}}\Xi), \\ tr(A_{\rho_\mu\varsigma_{(\nu\nu')}}\Xi), \quad tr(A^*_{\rho_\mu\rho_{\mu'}}\Xi), \quad tr(A^*_{\varsigma_{(\mu\mu')}\varsigma_{(\nu\nu')}}\Xi), \quad tr(A^*_{\rho_\mu\varsigma_{(\nu\nu')}}\Xi), \quad (A.68) \\ tr(A_{\rho_\mu} G A_{\rho_{\mu'}}\Xi), \quad tr(A_{\rho_\mu} G A_{\varsigma_{(\nu\nu')}}\Xi), \quad tr(A_{\varsigma_{(\mu\mu')}} G A_{\varsigma_{(\nu\nu')}}\Xi), \end{gathered}$$

with obvious modifications for

$$\begin{gathered} tr(A_{\varsigma_{(\nu\nu')}\rho_\mu}\Xi), \quad tr(A^*_{\varsigma_{(\nu\nu')}\rho_\mu}\Xi), \quad tr(A_{\varsigma_{(\nu\nu')}} G A_{\rho_\mu}\Xi), \\ tr(A_{\rho_\mu}\Xi A_{\rho_{\mu'}}\Xi), \quad tr(A_{\rho_\mu}\Xi A_{\varsigma_{(\nu\nu')}}\Xi), \quad tr(A_{\varsigma_{(\nu\nu')}}\Xi A_{\rho_\mu}\Xi), \quad tr(A_{\varsigma_{(\mu\mu')}}\Xi A_{\varsigma_{(\nu\nu')}}\Xi). \end{gathered}$$

By using the above results and Lemmas A.1 – A.14, the proof completes. □

Analytic formulae of the scalars and vectors given in (33) are derived in the following lemma.

Lemma A.17. *Scalars λ_0 and κ_0, vectors λ_ϱ, λ_ς, κ_ϱ and κ_ς, and matrices Λ_ϱ, Λ_ς and $\Lambda_{\varrho\varsigma}$ can be calculated as follows:*

$$\lambda_0 = 0, \quad \lambda_\varrho = 0, \quad \lambda_\varsigma = 0, \quad (A.69)$$

$$\Lambda_\varsigma = (\Sigma^{-1} \otimes \Sigma^{-1}) N (\Sigma^{-1} \otimes \Sigma^{-1}), \quad (A.70)$$

where N is a $(M^2 \times M^2)$ matrix whose $((ij),(kr))$-th element is

$$\nu_{(ij)(kr)} = \sigma_{ik}\sigma_{jr} + \sigma_{ir}\sigma_{jk} \quad (i,j,k,r = 1, \ldots, M). \quad (A.71)$$

The μ-th diagonal element of the matrix Λ_ϱ is

$$\lim_{T\to\infty} \mathcal{E}(d^2_{(1)\rho_\mu}) = 1 - \rho^2_\mu, \tag{A.72}$$

and its (μ, μ')-th off-diagonal element is

$$\lim_{T\to\infty} \mathcal{E}(d_{(1)\rho_\mu} d_{(1)\rho_{\mu'}}) = \frac{(1-\rho^2_\mu)(1-\rho^2_{\mu'})}{(1-\rho_\mu\rho_{\mu'})}, \tag{A.73}$$

for $\mu \neq \mu'$. Further, we have

$$\Lambda_{\varrho\varsigma} = 0 \quad and \quad \Lambda_{\varsigma\varrho} = 0. \tag{A.74}$$

For all estimators $\hat{\sigma}_I$ and $\hat{\varsigma}_I$ ($I = UL,\ RL,\ GL,\ IG,\ ML$), we can compute the following $(M \times M)$ matrices:

$$\Delta_{UL} = 0, \quad \Delta_{GL} = \Delta_{IG} = \Delta_{ML} = K\Sigma - \left[(tr(G_{ij}B_{ji}))_{i,j=1,\,\dots,\,M}\right],$$
$$\Delta_{RL} = \left[\left[\left(tr(B^{-1}_{ii}B_{ij}B^{-1}_{jj}B_{ji}) - n_i - n_j + K\right)\sigma_{ij}\right]_{i,j=1,\,\dots,\,M}\right]. \tag{A.75}$$

Given them we can calculate κ_0 and κ_ς as follows:

$$\kappa_0 = tr\left[\Sigma^{-1}(\Delta_{GL} - \Delta_I)\right]/M + n/M, \tag{A.76}$$

and

$$\kappa_\varsigma = vec\left\{(M+K+1)\Sigma^{-1} - \Sigma^{-1}\Delta_I\Sigma^{-1}\right\}. \tag{A.77}$$

Also, define scalars

$$c_1 = (1-\rho^2_\mu)[(1-\rho^2_\mu)tr(F^{-1}_{\mu\mu}\Theta_{\mu\mu}) + tr(F^{-1}_{\mu\mu}B_{\mu\mu}F^{-1}_{\mu\mu}\Theta_{\mu\mu})], \tag{A.78}$$

and

$$c_2 = (1-\rho^2_\mu)tr(F_{\mu\mu}B^{-1}_{\mu\mu}), \tag{A.79}$$

where

$$F_{\mu\mu} = X'_\mu X_\mu/T, \quad \Theta_{\mu\mu} = X'_\mu R_{\mu\mu} X_\mu/T \tag{A.80}$$

are $(n_\mu \times n_\mu)$ matrices. For all estimators $\hat{\rho}^I_\mu$ ($I = LS, GL, PW, ML,\ DW$), we calculate the elements κ_{ρ_μ} of $(M \times 1)$ vector κ_ϱ as follows:

$$\kappa^{LS}_{\rho_\mu} = -[(n_\mu + 3)\rho_\mu + (c_1 - 2n_\mu)/2\rho_\mu], \tag{A.81}$$

and

$$\begin{aligned}\kappa^{GL}_{\rho_\mu} = \kappa^{PW}_{\rho_\mu} &= \kappa^{LS}_{\rho_\mu} + \tfrac{c_1-(1-\rho^2_\mu)(c_2+n_\mu)}{2\rho_\mu},\\ \kappa^{ML}_{\rho_\mu} &= \kappa^{GL}_{\rho_\mu} + \rho_\mu, \\ \kappa^{DW}_{\rho_\mu} &= \kappa^{LS}_{\rho_\mu} + 1.\end{aligned} \tag{A.82}$$

Proof of Lemma A.17. From (33), (A.24), (A.29), and (A.48) we can easily show that

$$\lambda_0 = \lim_{T\to\infty} \mathcal{E}(\sigma_0^2), \quad \lambda_\rho = \lim_{T\to\infty} \mathcal{E}(\sigma_0 d_{1\rho}) \text{ and } \lambda_\varsigma = \lim_{T\to\infty} \mathcal{E}(\sigma_0 d_{1\varsigma}). \tag{A.83}$$

The results in (A.69) follows immediately since $\sigma_0 = 0$ (see(A.30)). Equations (33) and (A.24) imply

$$\Lambda_\varsigma = \lim_{T\to\infty} \mathcal{E}(d_{1\varsigma} d'_{1\varsigma}). \tag{A.84}$$

This result together with (A.25), (A.36) and (A.37) yield (A.70).

Since (33) and (A.48) imply that

$$\Lambda_\varrho = \lim_{T\to\infty} \mathcal{E}(d_{1\varrho} d'_{1\varrho}) \tag{A.85}$$

and $\sigma^2_{u_\mu} = \sigma^2_{\mu\mu}/(1-\rho^2_\mu)$, we can prove that the μ-th diagonal element of the matrix Λ_ϱ is

$$\begin{aligned}\lim_{T\to\infty} \mathcal{E}(d^2_{(1)\rho_\mu}) &= \lim_{T\to\infty} \mathcal{E}(u'_\mu R^{\mu\mu}_2 u_\mu u'_\mu R^{\mu\mu}_2 u_\mu)/4T\sigma^2_{u_\mu} \\ &= \lim_{T\to\infty} \left[\frac{4T\sigma^2_{\mu\mu}}{1-\rho^2_\mu} + O(1)\right]/4T\sigma^2_{u_\mu},\end{aligned} \tag{A.86}$$

by combining the third result in Lemma A.10 with (A.42) and (A.49). The last result proves (A.72). Working along the same lines for $\mu \neq \mu'$, we can prove (A.73), for the (μ, μ')-th off-diagonal element of Λ_ϱ.

To prove (A.74), first note that (33), (A.24) and (A.48) imply

$$\Lambda_{\varrho\varsigma} = \lim_{T\to\infty} \mathcal{E}(d_{1\varrho} d'_{1\varsigma}). \tag{A.87}$$

Substituting (A.25), (A.45) and (A.42) into (A.87), we can calculate the $(\mu, (ij))$-th element of $(M \times M^2)$ matrix $\Lambda_{\varrho\varsigma}$ as $-d_{(1)\rho_\mu} s^{(1)}_{(ij)}$. Following the same steps to that of the proof of (A.46) we can show that

$$\lim_{T\to\infty} \mathcal{E}\left(-d_{(1)\rho_\mu} s_{(ij)}^{(1)}\right) = 0. \tag{A.88}$$

(A.74) can be proved immediately using $\Lambda_{\varsigma\varrho} = \Lambda'_{\varrho\varsigma}$.

For all estimators $\hat{\sigma}_I$ ($I = UL$, RL, GL, IG, ML), we can find that

$$\kappa_0 = \lim_{T\to\infty} \mathcal{E}\left(\sqrt{T}\sigma_0 + \sigma_1\right) = \lim_{T\to\infty} \mathcal{E}\left(\sigma_1\right), \tag{A.89}$$

by combining (33) with (A.34), (A.29) and (A.30). The last result proves (A.76). For all estimators $\hat{\varsigma}_I$ ($I = UL$, RL, GL, IG, ML),we can show that

$$\kappa_\varsigma = \lim_{T\to\infty} \mathcal{E}\left(\sqrt{T}d_{1\varsigma} + d_{2\varsigma}\right) = vec\left\{\lim_{T\to\infty} \mathcal{E}\left(S_2^I\right)\right\}, \tag{A.90}$$

since $\mathcal{E}\left(S_1\right) = 0$ and $\lim_{T\to\infty} \mathcal{E}\left(S_2^I\right) = M_I$ [see (A.34)], by using (33), (A.24), (A.25) and (A.34). This result implies (A.77).

Finaly, we can calclulate

$$\kappa_{\rho_\mu}^{LS} = \lim_{T\to\infty} \mathcal{E}\left(\sqrt{T}d_{(1)\rho_\mu}^{LS} + d_{(2)\rho_\mu}^{LS}\right), \tag{A.91}$$

by using (33) and (A.42), Lemmas A.10 and A.13. This yields (A.81). Along the same lines, we can calculate the following quantities:

$$\begin{aligned} \kappa_{\rho_\mu}^{GL} = \kappa_{\rho_\mu}^{PW} &= \lim_{T\to\infty} \mathcal{E}\left(\sqrt{T}d_{(1)\rho_\mu}^{LS} + d_{(2)\rho_\mu}^{GL}\right), \\ \kappa_{\rho_\mu}^{ML} &= \lim_{T\to\infty} \mathcal{E}\left(\sqrt{T}d_{(1)\rho_\mu}^{LS} + d_{(2)\rho_\mu}^{ML}\right) \quad \text{and} \\ \kappa_{\rho_\mu}^{DW} &= \lim_{T\to\infty} \mathcal{E}\left(\sqrt{T}d_{(1)\rho_\mu}^{LS} + d_{(2)\rho_\mu}^{DW}\right), \end{aligned} \tag{A.92}$$

which proves (A.82). □

Asymptotic expansions of size corrected tests: Proofs of theorems

Given the lemmas of the previous subsections, next we give the proofs of the theorems presented in the main text. These are based on known expansions of standard normal and chi-square distributed tests. We derive new expansions of the degrees-of-freedom adjusted versions of these tests, by inverting their characteristic functions. These degrees-of-freedom adjusted are proved to be locally exact.

Proof of Theorems 1 and 2. Approximation (42) of Theorem 1 can be proved along the same lines with Rothenberg (1988). To obtain the quantities in (40), we expand the corresponding quantities given by Rothenberg and we retain the first term in the expansion. The approximation (44) of Theorem 2 follows from the approximation (42) and the following asymptotic approximations of the Student-t distribution and density functions. These are given in terms of the standard normal distribution and density functions, respectively (see Fisher (1925)):

$$I_{T-n}(x) = I(x) - (\tau^2/4)(1+x^2)xi(x) + O(\tau^4),$$

$$i_{T-n}(x) = i(x) + O(\tau^2). \tag{A.93}$$

Note that approximation (44) of Theorem 2 is locally exact. This can be easily seen as follows: If parameter vector $\gamma = (\varrho', \varsigma')'$ is known to belong to a ball of radius ϑ, then, as $\vartheta \to 0$, γ becomes a fixed known vector. By using (27), (29), (33) and (35) we can prove that

$$\Lambda = 0, \quad \lambda = \kappa = 0, \quad \lambda_0 = 2, \quad \kappa_0 = 0. \tag{A.94}$$

Then, the analytic formulae of p_1 and p_2, given in (43), become

$$p_1 = p_2 = 0. \tag{A.95}$$

This result implies that, with an error of order $O(\tau^3)$, approximation (44) becomes the Student-t distribution function with $MT - n$ degrees of freedom. □

Proof of Theorem 3. To prove the theorem, first notice that, under null hypothesis (36), the t statistic given by (37), admits a stochastic expansion of the form

$$t = t_0 + \tau t_1 + \tau^2 t_2 + \omega(\tau^3), \tag{A.96}$$

where the first term in the expansion is given as

$$t_0 = e'b/(e'Ge)^{1/2} = h'b, \quad \text{where} \quad b = GX'\Omega u/\sqrt{T}.$$

The result given by equation (A.96) implies that the Cornish-Fisher corrected statistic t_*, given by (47), admits a stochastic expansion of the form

$$t_* = t_0 + \tau t_1 + \tau^2(t_2 - t_3) + \omega(\tau^3), \tag{A.97}$$

where

$$t_3 = (p_1 + p_2 t_0^2)t_0/2.$$

given by (37) and a standard normal random variable, respectively. Using (A.97) and the relationships:

$$E[\exp(st_0)t_0] = s\phi(s) \quad \text{and} \quad E[\exp(st_0)t_0^3] = (3s + s^3)\phi(s),$$

we can show that the characteristic function of the Cornish-Fisher corrected statistic t_*, denoted as $\psi_*(s)$, can be approximated as follows:

$$\begin{aligned}\psi_*(s) &= \psi(s) - \tau^2 s\ E[\exp(st_0)t_3] + O(\tau^3) \\ &= \psi(s) - \frac{\tau^2}{2} s\ [p_1 s + p_2(3s + s^3)]\phi(s) + O(\tau^3).\end{aligned}$$

Dividing $\psi_*(s)$ by $-s$, applying the inverse Fourier transform and using Theorem 2, we can show that

$$\begin{aligned}\Pr\{t_* \leq x\} &= \Pr\{t \leq x\} + \frac{\tau^2}{2}(p_1 + p_2 x^2)x i_{T-n}(x) + O(\tau^3) \\ &= I_{T-n}(x) - \frac{\tau^2}{2}(p_1 + p_2 x^2)x i_{T-n}(x) \\ &\quad + \frac{\tau^2}{2}(p_1 + p_2 x^2)x i_{T-n}(x) + O(\tau^3) \\ &= I_{T-n}(x) + O(\tau^3). \end{aligned} \tag{A.98}$$

The last result means that the Cornish-Fisher corrected statistic t_* is distributed as a Student-t random variable with $MT - n$ degrees of freedom. □

Proof of Theorems 4 and 5. Approximation (58) of Theorem 4 can be proved along the same lines with Rothenberg (1984b). In order to obtain the quantities in (56), we expand the corresponding quantities given by Rothenberg and we retain the first term in the expansion. Approximation (60) of Theorem 5 follows from approximation (58) and the following asymptotic approximations of the F distribution and density functions. These are given in terms of the chi-square distribution and density functions, respectively:

$$F_{T-n}^{r}(x) = F_r(rx) + (\tau^2/2)(r-2-rx)rxf_r(rx) + O(\tau^4), \tag{A.99}$$

$$f_{T-n}^{r}(x) = rf_r(rx) + O(\tau^2).$$

Note that approximation (60) of Theorem 5 can be easily seen to be locally exact. By using (A.94), (59), and (61), we can show that

$$\xi_1 = -m(m-2)/2 \quad \text{and} \quad \xi_2 = m(m+2)/2 \tag{A.100}$$

$$\Rightarrow q_1 = q_2 = 0. \tag{A.101}$$

This result means that, with an error of order $O(\tau^3)$, approximation (60) becomes the F distribution function with m and $MT-n$ degrees of freedom. □

Proof of Theorem 6. To prove the theorem, first notice that, under null hypothesis (48), the F statistic given by (50) admits a stochastic expansion of the form

$$F = F_0 + \tau F_1 + \tau^2 F_2 + \omega(\tau^3), \tag{A.102}$$

where the first term in the expansion is

$$F_0 = b'Qb/r, \quad b = GX'\Omega u/\sqrt{T}.$$

Equation (A.102) implies that the Cornish-Fisher corrected statistic F_*, given by (64), admits a stochastic expansion of the form

$$F_* = F_0 + \tau F_1 + \tau^2(F_2 - F_3) + \omega(\tau^3), \tag{A.103}$$

where

$$F_3 = (q_1 + q_2 F_0) F_0.$$

Let s be an imaginary number, and $\psi(s)$ and $\phi_r(s)$ now denote the characteristic functions of the F statistic given by (50) and a chi-square random variable with r degrees of freedom, respectively. Using (A.103) and the following relationships:

$$E[\exp(sF_0)F_0] = \phi_{r+2}(s/r) \quad \text{and} \quad E[\exp(sF_0)F_0^2] = \frac{r+2}{r}\phi_{r+4}(s/r),$$

we can show that the characteristic function of the Cornish-Fisher corrected statistic F_*, denoted as $\psi_*(s)$, can be approximated as

$$\begin{aligned} \psi_*(s) &= \psi(s) - \tau^2 s \; E[\exp(sF_0)F_3] + O(\tau^3) \\ &= \psi(s) - \tau^2 s \; [q_1 \phi_{r+2}(s/r) + q_2 \frac{r+2}{r}\phi_{r+4}(s/r)] + O(\tau^3). \quad \text{(A.104)} \end{aligned}$$

For the chi-square density $f_r(x)$, the following results can be shown:

$$(rx)f_r(rx) = rf_{r+2}(rx) \quad \text{and} \quad (rx)^2 f_r(rx) = r(r+2)f_{r+4}(rx). \quad \text{(A.105)}$$

Dividing (A.104) by $-s$, applying the inverse Fourier transform, and using Theorem 5 and the results of equations (6) and (A.105), we can show that the following approximations hold:

$$\begin{aligned} \Pr\{F_* \le x\} &= \Pr\{F \le x\} + \tau^2[(q_1 r f_{r+2}(rx) + q_2 \frac{r+2}{r} r f_{r+4}(rx)] + O(\tau^3) \\ &= \Pr\{F \le x\} + \tau^2[(q_1 rx f_r(rx) + q_2 rx^2 f_r(rx)] + O(\tau^3) \\ &= \Pr\{F \le x\} + \tau^2(q_1 + q_2 x) rx f_r(rx) + O(\tau^3) \\ &= F^r_{T-n}(x) - \tau^2(q_1 + q_2 x) x f^r_{T-n}(x) \\ &\quad + \tau^2(q_1 + q_2 x) x f^r_{T-n}(x) + O(\tau^3) \\ &= F^r_{T-n}(x) + O(\tau^3). \quad \text{(A.106)} \end{aligned}$$

The last result implies that $F_* = F$, which means that the Cornish-Fisher corrected statistic F_* is distributed as an F random variable with m and $MT - n$ degrees of freedom. □

Bibliography

Beach, C. M. and MacKinnon, J. G. (1978). A Maximum Likelihood Procedure for Regression with Autocorrelated Errors. *Econometrica*, 46(1):51–58.

Breusch, T. S. (1980). Useful Invariance Results for Generalized Regression Models. *Journal of Econometrics*, 13:327–340.

Cornish, E. A. and Fisher, R. A. (1937). Moments and Cumulants in The Spesification of Distributions. *Revue del' Institute International de Statistic*, 4:1–14.

Dhrymes, P. J. (1971). Equivalence of Iterative Aitken and Maximum Likelihood Estimators for A System of Regression Equations. *Australian Economic Papers*, 10:20–24.

Dhrymes, P. J. (1978). *Introductory Econometrics*. Springer Verlang, New York.

Fisher, R. A. (1925). Expansion of Student's Integral in Powers of n^{-1}. *Metron*, 5:109–120.

Fisher, R. A. and Cornish, E. A. (1960). The Percentile of Distributions Having Known Cumulants. *Technometrics*, 2:209–225.

Hill, G. W. and Davis, A. W. (1968). Generalized Asymptotic Expansions of Cornish-Fisher Type. *The Annals of Mathematical Statistics*, 39(4):1264–1273.

Iglesias, E. M. and Phillips, G. D. (2010). The Bias to Order T^{-2} for the General k-class Estimator in A Simultaneous Equation Model. *Economics Letters*, 109:42–45.

Iglesias, E. M. and Phillips, G. D. (2011). Almost Unbiased Estimation in Simultaneus Equations Models with Strong and or Weak Instruments. Cardiff Economics Working Papers, No E2011/19.

Kiviet, J. and Phillips, G. D. (1996). The Bias of the Ordinary Least Squares Estimator in Simultaneous Equation Models. *Economics Letters*, (53):161–167.

Magdalinos, M. A. (1985). Improving Some Instrumental Variables Test Procedures. *Econometric Theory*, 1:240–261.

Magdalinos, M. A. (1992). Stochastic Expansions and Asymptotic Approximations. *Econometric Theory*, 8:343–367.

Magdalinos, M. A. and Symeonides, S. D. (1995). Alternative Size Corrections for Some GLS Test Statistics: The Case of the AR(1) Model. *Journal of Econometrics*, 66:35–59.

Magee, L. (1985). Efficiency of Iterative Estimators in the Regression Model with AR(1) Disturbances. *Journal of Econometrics*, 29:275–287.

Magnus, J. R. and Neudecker, H. (1979). The Commutation Matrix: Some Properties and Applications. *The Annals of Statistics*, 7:381–394.

McDonald, G. C. and Galarneau, D. I. (1975). A Monte Carlo Evaluation of Some Ridge-Type Estimators. *Journal of the American Statistical Association*, 70(350):407–416.

Parks, R. W. (1967). Efficient Estimation of a System of Regression Equations when Disturbances are Both Serially and Contemporaneously Correlated. *Journal of the American Statistical Association*, 62:500–509.

Phillips, G. D. (2000). An Alternative Approach to Obtaining Nagar-type Moment Approximations in Simultaneous Equation Models. *Journal of Econometrics*, 97:345–364.

Phillips, G. D. (2007). Nagar-type Moment Approximations in Simultaneous Equation Models: Some Further Results. In *The Refinement of Econometric Estimation and*

Test Procedures: Finite Sample and Asymptotic Analysis, pages 1–1. Cambridge University Press, Cambridge.

Prais, S. J. and Winsten, C. I. (1954). Trend Estimators and Serial Correlation. Cowles Commission Discussion Paper: Statistics No. 383.

Rothenberg, T. J. (1984a). Approximate Normality of Generalized Least Squares Estimates. *Econometrica*, 52(4):811–825.

Rothenberg, T. J. (1984b). Hypothesis Testing in Linear Models when the Error Covariance Matrix is Nonscalar. *Econometrica*, 52(4):827–842.

Rothenberg, T. J. (1988). Approximate Power Functions for Some Robust Tests of Regression Coefficients. *Econometrica*, 56(5):997–1019.

Srivastava, V. and Maekawa, K. (1995). Efficiency Properties of Feasible Generalized Least Squares Estimators in SURE Models under non-Normal Disturbances. *Journal of Econometrics*, 66:99–121.

Symeonides, S. D., Tzavalis, E., and Kandilorou, H. (2007). Cornish-Fisher Size Corrected t and F Statistics for the Linear Regression Model with Heteroskedastic Errors. In Philips, G. D. A. and Tzavalis, E., editors, *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, pages 173–204. Cambridge University Press, Cambridge.

Ullah, A. (2004). *Finite Sample Econometrics*. Oxford University Press.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association*, 57(298):348–368.

Zellner, A. (1963). Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results. *Journal of the American Statistical Association*, 58(304):977–992.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. John Wiley & Sons, New York.

Zellner, A. and Huang, D. S. (1962). Further Properties of Efficient Estimators for Seemingly Unrelated Regreesion Equations. *International Economic Review*, 3(3):300–313.

Zellner, A. and Theil, H. (1962). Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations. *Econometrica*, 30(1):54–78.