TECHNICAL WORKING PAPER SERIES

DYNAMIC SEEMINGLY UNRELATED COINTEGRATING REGRESSION

Nelson C. Mark
Masao Ogaki
Donggyu Sul

Technical Working Paper 292
http://www.nber.org/papers/T0292

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2003

We have benefited from comments on a previous draft by seminar participants at the University of California at Santa Barbara and the University of Southern California. We thank Ling Hu for a helpful discussion and two anonymous referees for useful comments. The views expressed in this paper are those of the authors and not necessarily those of the National Bureau of Economic Research.

Dynamic Seemingly Unrelated Cointegrating Regression
Nelson C. Mark, Masao Ogaki, and Donggyu Sul
NBER Technical Working Paper No. 292
April 2003
JEL No. C0, C3

ABSTRACT

Multiple cointegrating regressions are frequently encountered in empirical work as, for example, in the analysis of panel data. When the equilibrium errors are correlated across equations, the seemingly unrelated regression estimation strategy can be applied to cointegrating regressions to obtain asymptotically ecient estimators. While non-parametric methods for seemingly unrelated cointegrating regressions have been proposed in the literature, in practice, specification of the estimation problem is not always straightforward. We propose Dynamic Seemingly Unrelated Regression (DSUR) estimators which can be made fully parametric and are computationally straightforward to use. We study the asymptotic and small sample properties of the DSUR estimators both for heterogeneous and homogenous cointegrating vectors. The estimation techniques are then applied to analyze two long-standing problems in international economics. Our first application revisits the issue of whether the forward exchange rate is an unbiased predictor of the future spot rate. Our second application revisits the problem of estimating long-run correlations between national investment and national saving.

Nelson C. Mark
Department of Economics
Ohio State University
Columbus, OH 43210
and NBER
mark.1@osu.edu

Masao Ogaki
Department of Economics
Ohio State University
Columbus, OH 43210
Mogaki@ecolan.sbs.ohio-state.edu

Donggyu Sul
University of Auckland
d.sul@auckland.ac.nz

Introduction

Multiple-equation cointegrating regressions are frequently encountered in applied research. Many examples are found in the analysis of panel data. When the equilibrium errors are correlated across cross-sectional units, the idea of seemingly unrelated regressions (SUR) can be applied to cointegrating regressions to obtain asymptotically efficient estimators. Non-parametric methods for seemingly unrelated cointegrating regressions have previously been proposed by Park and Ogaki (1991), who applied the SUR method to generalize Park's (1992) Canonical Cointegrating Regression estimators and by Moon (1999) who applied the SUR method to generalize Phillips and Hansen's (1990) fully modified estimators.[1] The drawback of these SUR estimators, however, is that specification of the estimation problem is not always straightforward in practice. One particularly troublesome feature of these estimators is that the specific form of the non-parametric transformation that is required depends on the number of common regressors in the SUR equations.

In this paper, we propose Dynamic Seemingly Unrelated Regression (DSUR) Estimators for estimating small systems of cointegrating regressions. We study the asymptotic and small sample properties of the DSUR estimator which can be made fully parametric and are computationally straightforward to use. The methodology is feasible for balanced panels where N is substantially smaller than the number of time-series observations T. The asymptotic distribution theory that we use is for $T \to \infty$ and N fixed. We consider environments where the cointegrating vectors are homogeneous across equations and where they exhibit heterogeneity.

Cointegration vectors that exhibit cross-sectional heterogeneity can be estimated by DSUR or by dynamic ordinary least squares (DOLS) techniques. We compare DSUR to a generalized DOLS estimator developed by Saikkonen (1991) which, following the terminology of Park and Ogaki (1991), we call system DOLS. System DOLS is distinguished from ordinary DOLS proposed by Phillips and Loretan (1991), and Stock and Watson (1993) in that endogeneity in equation i is corrected by introducing leads and lags of the first difference not only of the regressors of equation i but also of the regressors of all other equations in the system. In the multivariate regression framework studied by Saikkonen (1991), the regressors are common in all regression equations. Therefore, there is no efficiency gain from the SUR method just as in the stationary case. Saikkonen (1991) shows that the system DOLS estimator is asymptotically efficient relative to the ordinary DOLS estimator in his framework.

In our framework, we allow different regressors to appear across the various cointe-

[1]After the first version of this paper was completed, we discovered that Moon and Perron (2000) also studied dynamic SUR.

grating regression equations. As in the stationary case the SUR method can be used to gain efficiency in our framework: the DSUR estimator achieves asymptotic efficiency gains over DOLS by incorporating the long-run cross-sectional correlation in the equilibrium errors in estimation. In addition, Wald statistics with limiting chi-square distributions can be conveniently constructed to test cross-equation restrictions—such as homogeneity restrictions—on the cointegration vectors. We also show that the computational burden can be lightened by focusing on the more convenient but asymptotically equivalent two-step DSUR estimator. In the first step, the regressand in each equation is regressed on the leads and lags of the first difference of the regressors from all equations to control for the endogeneity problem. In the second step, the SUR strategy is applied to the residuals from the first step regressions.

When the cointegration vector is homogenous across equations, estimation can be performed using a restricted version of the DSUR estimator. Restricted DSUR is is a pooled estimator of the cointegration vector that exploits the long-run dependence across individuals with the homogeneity restrictions across equations imposed in estimation. The comparison estimator under cointegration vector homogeneity is panel DOLS, which has previously been studied in the literature. Extant analyses of panel DOLS, however, have been conducted under the assumption of independence across cross-sectional units. We show below that under cross-sectional dependence, the asymptotic distribution of panel DOLS is straightforward to obtain.[2] Here as well, restricted DSUR achieves asymptotic efficiency gains relative to panel DOLS by incorporating the cross-equation dependence in the equilibrium errors in estimation.

In any finite sample, estimation of long-run covariance matrices can be a thorny task upon which estimator performance may hinge. It is therefore important to know whether or not the predictions from asymptotic theory are borne out in small samples. To address this question, we compare the small sample performance of alternative estimators in a series of Monte Carlo experiments. We find that the asymptotic distribution theory developed for all of the estimators work reasonably well and that there are important and sizable efficiency gains to be enjoyed by using DSUR over the DOLS methods.

We go on to illustrate the usefulness and computational feasibility of the DSUR method by revisiting two long-standing problems in international economics. The first application revisits Evans and Lewis's (1995) cointegrating regressions of the future spot exchange rate on the current forward exchange rate which asks whether the forward rate is an unbiased predictor of the future spot rate. Using ordinary DOLS, they report a new

[2] Mark and Sul (1999) and Kao and Chiang (1998) studied the properties of panel DOLS under the assumption of independence across cross-sectional units. Pedroni (1997) and Phillips and Moon (1998) study a panel fully modified OLS estimator also under cross-sectional independence. Moreover, the asymptotic theory employed in these papers requires both T and N to go to infinity.

anomaly in international finance—that the slope coefficient is significantly different from 1—from which it follows that the expected excess return from forward foreign exchange speculation is unit-root nonstationary. When we update Evans and Lewis's sample and employ DSUR, we find the evidence for a nonstationary expected excess return to be less compelling.

Our second application revisits the estimation of national saving and investment correlations put forth by Feldstein and Horioka (1980). Their interpretation is that the size of the estimated slope coefficient in a regression of the national investment to GDP ratio on the national saving to GDP ratio is inversely related to the degree of capital mobility. Feldstein and Horioka found that the slope coefficient in their regression was insignificantly different from 1, from which they conclude that the degree of international capital mobility is low. The original Feldstein–Horioka analysis employed a cross-sectional regression using time-series averages as observations. Coakley et al. (1996) extend this work to the time-series dimension. These authors show that under a time-series interpretation, a solvency constraint restricts the current account balance to be stationary irrespective of the degree of capital mobility. Because the current account is saving minus investment, it is possible that Feldstein and Horioka's cross-section regression may just be capturing this long-run relationship when long-run time series averages are used for the regression. In our panel data application, we regress investment variables onto saving variables as a system of cointegrating regressions and test the hypothesis that the slope coefficient is 1. This provides a more direct test of the long-run relationship implied by the solvency condition than cross-section regressions.

The remainder of the paper is organized as follows. The next section presents and discusses the asymptotic properties of the alternative estimators that we examine. In section 2 we conduct a Monte Carlo experiment to examine the small sample performance of estimators and the accuracy of the asymptotic approximations. In section 3 we apply the estimators to the spot–forward exchange rate problem and to the investment–saving puzzle. Section 4 concludes the paper. Proofs of propositions are contained in the appendix.

1 System Estimators of Cointegration Vectors

We consider N cointegrating regressions where N is fixed. The data are balanced panels of individuals indexed by $i = 1, \ldots, N$ tracked over time periods $t = 1, \ldots, T$. Our notational conventions are as follows: Vectors are underlined and matrices appear in bold face but scalars have no special notation. $\underline{W}(r)$ is a vector standard Brownian motion for $0 \leq r \leq 1$, and $[Tr]$ denotes the largest integer value of Tr for $0 \leq r \leq 1$. We will not make the notational dependence on r explicit, so integrals such as $\int_0^1 \underline{W}(r)dr$

are written as $\int \underline{W}$ and $\int_0^1 \underline{W}(r)d\underline{W}(r)'$ are written as $\int \underline{W}d\underline{W}'$. Scaled vector Brownian motions are denoted by $\underline{B} = \mathbf{\Lambda}\underline{W}$ where $\mathbf{\Lambda}$ is a scaling matrix. The regularity conditions that we impose are given in,

Assumption 1 (Triangular Representation.) Each equation has the triangular representation,

$$y_{it} = \underline{x}_{it}'\underline{\beta}_i + u_{it}^\dagger, \tag{1}$$

$$\Delta\underline{x}_{it} = \underline{e}_{it}, \tag{2}$$

where $\underline{x}_{it}$ and $\underline{e}_{it}$ are $k\times 1$-dimensional vectors, $\underline{w}_t^\dagger = (\underline{u}_t^{\dagger'}, \underline{e}_t')'$ is an $N(k+1)$-dimensional vector with the orthonomal Wold moving average representation, $\underline{w}_t^\dagger = \mathbf{\Psi}^\dagger(L)\underline{\epsilon}_t$, in which $\underline{\epsilon}_t$ is serially uncorrelated with $\mathrm{E}(\underline{\epsilon}_t) = \underline{0}$, $\mathrm{E}(\underline{\epsilon}_t\underline{\epsilon}_t') = \mathbf{I}_k$, $\underline{u}_t^\dagger = (u_{1t}^\dagger, \ldots, u_{NT}^\dagger)'$, $\underline{e}_t' = (\underline{e}_{1t}', \ldots, \underline{e}_{NT}')'$, $\sum_{r=0}^\infty r|\psi_{ir}^{mn}| < \infty$, and ψ_{ir}^{mn} is the $m,n-$th element of the matrix $\mathbf{\Psi}_{ir}$.

It follows from Assumption 1 that $\underline{w}_t^\dagger$ obeys the functional central limit theorem, $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} \underline{w}_t^\dagger \xrightarrow{D} \underline{B}^\dagger(r) = \mathbf{\Psi}^\dagger(1)\underline{W}(r)$ where $\underline{B}^\dagger = (\underline{B}_u^\dagger, \underline{B}_{e_1}', \ldots, \underline{B}_{e_N}')'$ is an $N(k+1)$–dimensional scaled vector Brownian motion with covariance matrix, $\mathbf{\Omega}^\dagger = \mathbf{\Psi}^\dagger(1)\mathbf{\Psi}^\dagger(1)' = \sum_{j=-\infty}^\infty \mathrm{E}[\underline{w}_j^\dagger \underline{w}_0^{\dagger'}]$
$= \mathbf{\Gamma}_0^\dagger + \sum_{j=1}^\infty (\mathbf{\Gamma}_j^\dagger + \mathbf{\Gamma}_j^{\dagger'})$. The long-run covariance matrix and its components can be partitioned as,

$$\mathbf{\Omega}^\dagger = \begin{bmatrix} \mathbf{\Omega}_{uu}^\dagger & \mathbf{\Omega}_{ue}^\dagger \\ \mathbf{\Omega}_{eu}^\dagger & \mathbf{\Omega}_{ee}^\dagger \end{bmatrix} = \begin{bmatrix} \mathbf{\Omega}_{uu}^\dagger & \mathbf{\Omega}_{ue_1}^\dagger & \cdots & \mathbf{\Omega}_{ue_N}^\dagger \\ \mathbf{\Omega}_{e_1u}^\dagger & \mathbf{\Omega}_{e_1e_1}^\dagger & \cdots & \mathbf{\Omega}_{e_1e_N}^\dagger \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Omega}_{e_Nu}^\dagger & \mathbf{\Omega}_{e_Ne_1}^\dagger & \cdots & \mathbf{\Omega}_{e_Ne_N}^\dagger \end{bmatrix},$$

$$\mathbf{\Gamma}_j^\dagger = \begin{bmatrix} \mathbf{\Gamma}_{uu,j}^\dagger & \mathbf{\Gamma}_{ue,j}^\dagger \\ \mathbf{\Gamma}_{eu,j}^\dagger & \mathbf{\Gamma}_{ee,j}^\dagger \end{bmatrix} = \begin{bmatrix} \mathbf{\Gamma}_{uu,j}^\dagger & \mathbf{\Gamma}_{ue_1,j}^\dagger & \cdots & \mathbf{\Gamma}_{ue_N,j}^\dagger \\ \mathbf{\Gamma}_{e_1u,j}^\dagger & \mathbf{\Gamma}_{e_1e_1,j}^\dagger & \cdots & \mathbf{\Gamma}_{e_1e_N,j}^\dagger \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Gamma}_{e_Nu,j}^\dagger & \mathbf{\Gamma}_{e_Ne_1,j}^\dagger & \cdots & \mathbf{\Gamma}_{e_Ne_N,j}^\dagger \end{bmatrix}$$

where $\mathbf{\Gamma}_{uu,j}^\dagger = \mathrm{E}(\underline{u}_t^\dagger \underline{u}_{t-j}^{\dagger'})$, $\mathbf{\Gamma}_{ue_k,j}^\dagger = \mathrm{E}(\underline{u}_t^\dagger \underline{e}_{kt-j}')$, and $\mathbf{\Gamma}_{e_ke_s,j}^\dagger = \mathrm{E}(\underline{e}_{kt}\underline{e}_{st-j}')$.

Because $\mathbf{\Omega}_{e_iu}^\dagger$ is the long-run covariance between $\underline{e}_{it}$ and $(u_{1t}^\dagger, \ldots, u_{Nt}^\dagger)$, $i = 1, \ldots, N$, the endogeneity problem shows up as correlation between the equilibrium error of equation i and leads and lags of first differences of the regressors of all of the other equations

$j = 1, \ldots, N$. In system estimation methods, parametric adjustments for endogeneity in equation $i = 1$ will in general require inclusion of leads and lags not only of $\Delta \underline{x}_{1t}$, as is the case in the single-equation environment or in the panel environment under cross-sectional independence, but also leads and lags of $\Delta \underline{x}_{2t}$ through $\Delta \underline{x}_{Nt}$ as well.

The next subsection discusses estimation strategies for heterogeneous cointegration vectors. Section 1.2 discusses estimation of a homogeneous cointegration vector.

1.1 Estimation of Heterogeneous Cointegration Vectors

The asymptotic distributions that we derive are obtained by letting $T \to \infty$ for fixed N. For concreteness and without loss of generality, we set $N = 2$. Section 2.2 introduces and discusses the properties of the DSUR estimator. An asymptotically equivalent but computationally more convenient two-step DSUR estimator is discussed in section 1.1.2. In section 1.1.3, we discuss the joint distribution of system DOLS.

1.1.1 DSUR

$u_{it}^{\dagger}$ is potentially correlated with all leads and lags of $\Delta \underline{x}_{jt} = \underline{e}_{jt}$, $(i, j = 1, 2)$. In any feasible parametric estimation strategy only a finite number p of leads and lags can be included so in general, a cutoff at p will induce a separate truncation error. To keep track of the truncation error, let

$$\begin{aligned}
\underline{z}'_{pit} &= (\Delta \underline{x}'_{it-p}, \ldots, \Delta \underline{x}'_{it+p}), \\
\underline{z}'_{pt} &= (\underline{z}'_{p1t}, \underline{z}'_{p2t}), \\
\underline{\delta}_{p1} &= \left(\underline{\delta}'_{11,-p}, \ldots \underline{\delta}'_{11,p}, \underline{\delta}'_{12,-p} \ldots, \underline{\delta}'_{12,p}\right), \quad \text{and} \\
\underline{\delta}_{p2} &= \left(\underline{\delta}'_{21,-p}, \ldots \underline{\delta}'_{21,p}, \underline{\delta}'_{22,-p} \ldots, \underline{\delta}'_{22,p}\right),
\end{aligned}$$

where $\underline{\delta}_{ij,p}$ is a $k \times 1$ vector of coefficients. Under the conditions of Assumption 1, the equilibrium errors can be represented as

$$u_{1t}^{\dagger} = \underline{z}'_{pt}\underline{\delta}_{p1} + v_{p1t} + u_{1t}, \tag{3}$$
$$u_{2t}^{\dagger} = \underline{z}'_{pt}\underline{\delta}_{p2} + v_{p2t} + u_{2t}, \tag{4}$$

where

$$v_{p1t} = \sum_{j>|p|} \underline{\delta}'_{11,j} \Delta \underline{x}_{1t-j} + \sum_{j>|p|} \underline{\delta}'_{12,j} \Delta \underline{x}_{2,t-j}, \tag{5}$$
$$v_{p2t} = \sum_{j>|p|} \underline{\delta}'_{21,j} \Delta \underline{x}_{1t-j} + \sum_{j>|p|} \underline{\delta}'_{22,j} \Delta \underline{x}_{2,t-j}, \tag{6}$$

are the truncation errors induced for given p arising from the dependence of the equilibrium errors on $(\Delta \underline{x}'_{1t}, \Delta \underline{x}'_{2t})$ at distant leads and lags. Substituting (3) and (4) into (1) yields the regression $y_{it} = \underline{x}'_{it}\underline{\beta}_i + \underline{z}'_{pt}\underline{\delta}_{pi} + v_{pit} + u_{it}$. If we let $\underline{y}_t = (y_{1t}, y_{2t})'$, $\underline{u}_t = (u_{1t}, u_{2t})'$, $\underline{v}_{pt} = (v_{p1t}, v_{p2t})'$, $\underline{\beta} = (\underline{\beta}'_1, \underline{\beta}_2)'$, $\underline{\delta}_p = (\underline{\delta}'_{p1}, \underline{\delta}'_{p2})'$, $\mathbf{Z}_{pt} = (\mathbf{I}_2 \otimes \underline{z}_{pt})$, $\mathbf{X}_t = \text{diag}\,(\underline{x}_{1t}, \underline{x}_{2t})$ and $\mathbf{W}_t = \left(\mathbf{X}'_t, \mathbf{Z}'_{pt}\right)'$, the equations can be stacked together in a system as,

$$\underline{y}_t = \left(\underline{\beta}', \underline{\delta}'_p\right) \mathbf{W}_t + \underline{v}_{pt} + \underline{u}_t. \tag{7}$$

The DSUR estimator with known $\mathbf{\Omega}_{uu}$ is,

$$\begin{bmatrix} \underline{\hat{\beta}}_{dsur} \\ \underline{\hat{\delta}}_{p,dsur} \end{bmatrix} = \left(\sum_{t=p+1}^{T-p} \mathbf{W}_t \mathbf{\Omega}_{uu}^{-1} \mathbf{W}'_t\right)^{-1} \left(\sum_{t=p+1}^{T-p} \mathbf{W}_t \mathbf{\Omega}_{uu}^{-1} \underline{y}_t\right). \tag{8}$$

Due to the stationarity of the equilibrium errors, the dependence of $u^{\dagger}_{it}$ on $\Delta \underline{x}_{jt}$ at very distant leads and lags becomes trivial. Under the regularity conditions of Saikkonen (1991) it can be shown that by allowing the number of leads and lags of changes in the regressors to increase at a certain rate with T, the truncation errors will vanish asymptotically. We follow Saikkonen in

Assumption 2 (Lead and lag dependence.) Let $p(T)$ be the number of leads and lags of $\Delta \underline{x}_{it}$, $(i = 1, 2)$ included in the regression (7). We assume that

i. $p(T)/T^{1/3} \to 0$ as $T \to \infty$, and

ii. $\sqrt{T} \sum_{|j|>p(T)} \left\| \begin{matrix} \underline{\delta}'_{11,j} & \underline{\delta}'_{12,j} \\ \underline{\delta}'_{21,j} & \underline{\delta}'_{22,j} \end{matrix} \right\| \to 0,$

where $||\cdot||$ is the Euclidian norm.

The second condition in Assumption 2 places an upper bound on the allowable dependence of $u^{\dagger}_{it}$ on $\Delta \underline{x}_{jt}$ at very distant leads and lags, while the first condition controls the rate at which additional leads and lags must be included in order for the truncation induced misspecification error to vanish. We are now ready to state our first result.

Proposition 1 (Asymptotic distribution of DSUR). Let $T_* = T - 2p$. Under the conditions of Assumptions 1 and 2,

a. $T_* \left(\hat{\underline{\beta}}_{dsur} - \underline{\beta}\right)$ and $\sqrt{T_*} \left(\hat{\underline{\delta}}_{p,dsur} - \underline{\delta}_p\right)$ are asymptotically independent.

b. If $\mathbf{B}_e = \text{diag}\left(\underline{B}_{e_1}, \underline{B}_{e_2}\right)$, $\hat{\mathbf{V}}_{dsur} = \sum_{t=p+1}^{T-p} \mathbf{X}_t \mathbf{\Omega}_{uu}^{-1} \mathbf{X}_t'$, and $\mathbf{R}$ is a $q \times 2k$ matrix of constants such that $\mathbf{R}\underline{\beta} = \underline{r}$, then as $T_* \to \infty$,

$$T_*(\hat{\underline{\beta}}_{dsur} - \underline{\beta}) \xrightarrow{D} \left(\int \mathbf{B}_e \mathbf{\Omega}_{uu}^{-1} \mathbf{B}_e'\right)^{-1} \int \mathbf{B}_e \mathbf{\Omega}_{uu}^{-1} d\underline{B}_u, \tag{9}$$

and

$$(\mathbf{R}\hat{\underline{\beta}}_{dsur} - \underline{r})' \left[\mathbf{R}\hat{\mathbf{V}}_{dsur}\mathbf{R}'\right]^{-1} (\mathbf{R}\hat{\underline{\beta}}_{dsur} - \underline{r}) \xrightarrow{D} \chi_q^2. \tag{10}$$

The intuition behind Proposition 1 is that asymptotically, as the effects of the truncation error become trivial, one obtains a newly defined vector process $\underline{w}_t' = (u_{1t}, u_{2t}, \underline{e}_{1t}', \underline{e}_{2t}')$, with the moving average representation,

$$\underline{w}_t = \begin{bmatrix} \mathbf{\Psi}_{11}(L) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Psi}_{22}(L) \end{bmatrix} \begin{bmatrix} \underline{v}_{1t} \\ \underline{v}_{2t} \end{bmatrix},$$

where $\mathbf{\Psi}_{11}(L)$ and $\mathbf{\Psi}_{22}(L)$ are (2×2) and $(2k\times2k)$ matrix polynomials in the lag operator L, respectively, and which obeys the functional central limit theorem, $\frac{1}{\sqrt{T_*}} \sum_{t=p+1}^{[(T-p)r]} \underline{w}_t \xrightarrow{D} (\underline{B}_u', \underline{B}_e')'$ with long-run covariance matrix, $\mathbf{\Omega} = \text{diag}\left(\mathbf{\Omega}_{uu}, \mathbf{\Omega}_{ee}\right)$. By the block diagonality of $\mathbf{\Omega}$, it is seen that $\underline{B}_u$ and $\underline{B}_e$ are independent.

In applications, we replace $\mathbf{\Omega}_{uu}$ with a consistent estimator, $\hat{\mathbf{\Omega}}_{uu} \xrightarrow{p} \mathbf{\Omega}_{uu}$. Such an estimator might be called a 'feasible' DSUR estimator. It is easy to see that the asymptotic distribution of the feasible DSUR estimator is identical to the DSUR estimator of Proposition 1. Accordingly, we will in general not make a distinction between estimators formed with a known $\mathbf{\Omega}_{uu}$ or one that is estimated.

Finally, we note that the Wald statistic defined in (10) provides a convenient test of homogeneity restrictions on the cointegrating vectors, $H_0 : \beta_1 = \beta_2$.

1.1.2 Two-step DSUR

Some computational economies can be achieved by conducting estimation in two steps. The first step purges endogeneity by least squares and the second step estimates $\underline{\beta}$ by running SUR on the least squares residuals obtained from the first-step regressions. This procedure is asymptotically equivalent to the one-step DSUR estimator discussed above. When the number p of included leads and lags are identical across equations, this OLS-SUR two-step estimator is numerically equivalent to a two-step procedure in which endogeneity is purged by generalized least squares (GLS) in the first step and then running SUR on these GLS residuals.

To form the two-step estimator, let $\underline{z}'_{pt}\hat{\underline{\gamma}}^{y}_{pi}$ be the fitted least-squares regression of y_{it} onto $\underline{z}_{pt}$ and let $(\mathbf{I}_k \otimes \underline{z}'_{pt})\hat{\underline{\gamma}}^{x}_{pi}$ be the vector of fitted least-squares regressions of $\underline{x}_{it}$ onto $\underline{z}'_{pt}$. Denote the regression errors by $\hat{y}_{it} = y_{it} - \underline{z}'_{pt}\hat{\underline{\gamma}}^{y}_{pi}$, and $\hat{\underline{x}}_{it} = \underline{x}_{it} - (\mathbf{I}_k \otimes \underline{z}'_{pt})\hat{\underline{\gamma}}^{x}_{pi}$. We can now represent the equation system as $\hat{y}_{it} = \hat{\underline{x}}'_{it}\underline{\beta}_i + \hat{u}_{it}$, where

$$
\begin{aligned}
\hat{u}_{it} &= \underline{z}'_{pt}(\underline{\delta}_{pi} - \hat{\gamma}^{y}_{pi}) + \left[(\mathbf{I}_k \otimes \underline{z}'_{pt})\hat{\underline{\gamma}}^{x}_{pi}\right]\underline{\beta}_i + u_{it} \\
&= \underline{z}'_{pt}\left(\underline{\delta}_{pi} - \hat{\underline{\delta}}_{pi,ols}\right) + u_{it},
\end{aligned}
$$

and $\hat{\underline{\delta}}_{pi,ols} = \hat{\underline{\gamma}}^{y}_{pi} - \underline{\beta}'_i\hat{\underline{\gamma}}^{x}_{pi}$. Now stacking the equations together in the system gives $\hat{\underline{y}}_t = \hat{\mathbf{X}}'_t\underline{\beta} + \hat{\underline{u}}_t$. The two-step DSUR estimator is

$$
\hat{\underline{\beta}}_{2sdsur} = \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{X}}_t\mathbf{\Omega}^{-1}_{uu}\hat{\mathbf{X}}'_t\right]^{-1}\left[\sum_{t=p+1}^{T-p} \hat{\mathbf{X}}_t\mathbf{\Omega}^{-1}_{uu}\hat{\underline{y}}_t\right], \tag{11}
$$

and its properties are given in

Proposition 2 (Asymptotic equivalence of the two-step estimator.) Under the conditions of Assumptions 1 and 2, the two-step DSUR estimator (11) is asymptotically equivalent to the one-step DSUR estimator of proposition 1. Moreover, if the same set of leads and lags $\underline{z}_{pt}$ is included in every equation, this OLS-SUR two-step estimator is numerically equivalent to a two-step estimator where endogeneity is purged by GLS and running SUR on the GLS residuals.

The asymptotic equivalence obtains due to the consistency of $\hat{\underline{\delta}}_{pi,ols}$ and its asymptotic independence of the estimator of $\underline{\beta}$. Since asymptotic equivalence is achieved in regressions using least squares residuals from first-step regressions, we will henceforth assume that endogeneity has been controlled for in this fashion and will work in terms of these first-step regression residuals.

1.1.3 DOLS

DOLS is a single-equation estimator and may ignore dependence across individuals in estimation. Controlling for endogeneity in equation i can be achieved by projecting $u^{\dagger}_{it}$

onto $\underline{z}_{pit}$ or onto $\underline{z}_{pt} = (\underline{z}'_{p1t}, \underline{z}'_{p2t})'$ as in DSUR. The first option involves only those time series that explicitly appear in equation i and is a member of what Saikkonen (1991) calls the $\mathcal{S}_2$ class. The second option, which employs auxiliary observations, is an example of what he calls the $\mathcal{S}_C$ class. Park and Ogaki (1991) consider a similar distinction in their study of canonical cointegrating regressions (CCR). We conform to Park and Ogaki's terminology and refer to the procedure that controls for endogeneity by conditioning on $\underline{z}_{pt}$ as the 'system' DOLS estimator. We call the estimator that conditions on $\underline{z}_{pit}$ 'ordinary' DOLS.

While the joint distribution of DOLS across equations depends on the long-run covariance matrix, $\boldsymbol{\Omega}_{uu}$, the estimator itself does not exploit this information. Here, we discuss two-step estimation of system DOLS and compare it to DSUR. In two-step system DOLS, endogeneity can be purged by least squares and then and then the cointegration vector estimated by running OLS on the residuals from the first-step regressions.

Let $\hat{y}_{it}$ be the error obtained from regressing y_{it} on $\underline{z}_{pt}$ and let $\hat{\underline{x}}_{it}$ be the $k \times 1$ vector of errors obtained from regressing each element of $\underline{x}_{it}$ on $\underline{z}_{pt}$. Stacking the equations together as the system gives $\underline{\hat{y}}_t = \hat{\mathbf{X}}'_t \underline{\beta} + \underline{u}_t$, where the dimensionality of the matrices are as defined above. The system DOLS estimator is[3]

$$\underline{\hat{\beta}}_{sysdols} = \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{X}}_t \hat{\mathbf{X}}'_t\right]^{-1} \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{X}}_t \underline{\hat{y}}_t\right], \tag{12}$$

for which we have,

Proposition 3 (Asymptotic distribution of system DOLS). *Under the conditions of Assumptions 1 and 2, as $T_* \to \infty$,*

$$T_*(\underline{\hat{\beta}}_{sysdols} - \underline{\beta}) \xrightarrow{D} \left(\int \mathbf{B}_e \mathbf{B}'_e\right)^{-1} \left(\int \mathbf{B}_e d\underline{B}_u\right) = \begin{pmatrix} \left(\int \underline{B}_{e_1}\underline{B}'_{e_1}\right)^{-1} \int \underline{B}_{e_1} d\underline{B}_{u_1} \\ \left(\int \underline{B}_{e_2}\underline{B}'_{e_2}\right)^{-1} \int \underline{B}_{e_2} d\underline{B}_{u_2} \end{pmatrix}, \tag{13}$$

---

[3]If we let $\mathbf{X}_T = \begin{bmatrix} \hat{\underline{x}}'_{1p+1} & \underline{0}' \\ \vdots & \vdots \\ \hat{\underline{x}}'_{1T-p} & \underline{0}' \\ \underline{0}' & \hat{\underline{x}}'_{2p+1} \\ \vdots & \vdots \\ \underline{0}' & \hat{\underline{x}}'_{2T-p} \end{bmatrix}$, then in the standard matrix notation, $\mathbf{V}_{sysdols} = (\mathbf{X}'_T \mathbf{X}_T)^{-1} \mathbf{X}'_T (\boldsymbol{\Omega}_{uu} \otimes \mathbf{I}_T) \mathbf{X}_T (\mathbf{X}'_T \mathbf{X}_T)^{-1}$.

and

$$(\mathbf{R}\hat{\underline{\beta}}_{sysdols} - \underline{r})' \left[\mathbf{R}\hat{\mathbf{V}}_{sysdols}\mathbf{R}'\right]^{-1} (\mathbf{R}\hat{\underline{\beta}}_{sysdols} - \underline{r}) \xrightarrow{D} \chi^2_q, \tag{14}$$

where $\hat{\mathbf{V}}_{sysdols} = \left[\sum_{t=p+1}^{T-p} \mathbf{X}_t\mathbf{X}_t'\right]^{-1} \left[\sum_{t=p+1}^{T-p} \mathbf{X}_t\mathbf{\Omega}_{uu}\mathbf{X}_t'\right] \left[\sum_{t=p+1}^{T-p} \mathbf{X}_t\mathbf{X}_t'\right]^{-1}$ and $\mathbf{R}$ is a $q \times 2k$ matrix of constants such that $\mathbf{R}\underline{\beta} = \underline{r}$.

Saikkonen showed that within the context of the standard multivariate regression framework, ordinary DOLS is efficient within the class of $\mathcal{S}_2$ estimators and that the class of $\mathcal{S}_C$ estimators are efficient relative to the $\mathcal{S}_2$ class. The reason for this is as follows. In ordinary DOLS, endogeneity is purged by projecting $u_{it}^{\dagger}$ onto $\underline{z}_{pit}$. Substituting this projection representation into (1) gives $y_{it} = \underline{x}_{it}'\underline{\beta}_i + \underline{\lambda}_i'\underline{z}_{it}' + \zeta_{it}$, where ζ_{it} is the projection error which is by construction orthogonal to included leads and lags of $\Delta\underline{x}_{it}$. Since $(1/\sqrt{T})\sum(\zeta_{it}, \underline{e}_{it}')' \xrightarrow{D} (B_{\zeta i}, \underline{B}_{ei}')'$ with long-run covariance matrix diag$(\Omega_{\zeta_i,\zeta_i}, \mathbf{\Omega}_{e_i,e_i})$, it follows that conditional on $\underline{B}_{e_i}$, avar$(\hat{\underline{\beta}}_{dols}) = \Omega_{\zeta_i,\zeta_i}\left(\int \underline{B}_{e_i}\underline{B}_{e_i}'\right)^{-1}$. Since Ω_{ζ_i,ζ_i} is the long-run variance of the error from projecting $u_{it}^{\dagger}$ onto $\underline{z}_{pit} \subseteq \underline{z}_{pt}$ and Ω_{u_i,u_i} is the long-run variance of the error from projecting $u_{it}^{\dagger}$ onto $\underline{z}_{pt}$, it must be the case that $\Omega_{\zeta_i,\zeta_i} \geq \Omega_{u_i,u_i}$. Thus, avar$(\hat{\underline{\beta}}_{i,dols}) \geq$ avar$(\hat{\underline{\beta}}_{i,sysdols})$.

Our representation of the observations (Assumption 1) differs from Saikkonen's in that it imposes 'zero-restrictions' on the multivariate regression in whereby each equation contains a different set of regressors. Thus in the context of the model that we study, DSUR, which exploits the cross-equation correlations, enjoys asymptotic efficiency advantages over single-equation methods. A comparison of the asymptotic efficiency of system DOLS and DSUR gives

Proposition 4 Under the conditions of Assumptions 1 and 2, avar$(\hat{\underline{\beta}}_{dsur}) \leq$ avar$(\hat{\underline{\beta}}_{sysdols})$.

1.2 Estimation of Homogeneous Cointegration Vectors

We now turn to estimation of the cointegration vector under homogeneity, $\underline{\beta}_1 = \underline{\beta}_2 = \underline{\beta}$. We first discuss the restricted DSUR estimator. This is the DSUR estimator discussed above with homogeneity restrictions imposed and has a generalized least squares interpretation. In section 1.2.2, restricted DSUR is compared to the panel DOLS estimator.

1.2.1 Restricted DSUR

As in two-step DSUR, endogeneity can first be purged by regressing y_{it} and each element of $\underline{x}_{it}$ on $\underline{z}_{pt}$. Let $\hat{y}_{it}$ and $\hat{\underline{x}}_{it}$ denote the resulting regression errors. The problem is to estimate $\underline{\beta}$, in the system of equations $\hat{y}_{it} = \hat{\underline{x}}_{it}'\underline{\beta} + \hat{u}_{it}$ where $\underline{\beta}_1 = \underline{\beta}_2 = \underline{\beta}$. Stacking these equations together, we have,

$$\hat{\underline{y}}_t = \hat{\mathbf{x}}_t'\underline{\beta} + \hat{\underline{u}}_t \tag{15}$$

where $\hat{\mathbf{x}}_t = (\underline{x}_{1t}, \underline{x}_{2t})$ is a $k \times 2$ matrix.

Let $\mathbf{\Omega}_{uu} = \mathbf{L}\mathbf{L}'$ be the lower-triangular Choleski decomposition of the long-run error covariance matrix, where $\mathbf{L} = \begin{bmatrix} \ell_{11} & 0 \\ \ell_{21} & \ell_{22} \end{bmatrix}$, $\mathbf{L}^{-1} = \begin{bmatrix} \ell^{11} & 0 \\ \ell^{21} & \ell^{22} \end{bmatrix}$, $\ell^{11} = 1/\ell_{11}$, $\ell^{22} = 1/\ell_{22}$, and $\ell^{21} = -\ell_{21}/(\ell_{11}\ell_{22})$. We pre-multiply (15) by $\mathbf{L}^{-1}$ to get, $\hat{\underline{y}}_t^* = \hat{\mathbf{x}}_t^{*\prime}\underline{\beta} + \hat{\underline{u}}_t^*$ where $\hat{\underline{y}}_t^* = \mathbf{L}^{-1}\hat{\underline{y}}_t$, $\hat{\mathbf{x}}_t^* = \hat{\mathbf{x}}_t(\mathbf{L}^{-1})'$, and $\hat{\underline{u}}_t^* = \mathbf{L}^{-1}\hat{\underline{u}}_t$. The restricted DSUR estimator is obtained by running OLS on these transformed observations,

$$\hat{\underline{\beta}}_{rdsur} = \left[\sum_{i=1}^{2}\sum_{t=p+1}^{T-p} \hat{\underline{x}}_{it}^*\hat{\underline{x}}_{it}^{*\prime}\right]^{-1} \left[\sum_{i=1}^{2}\sum_{t=p+1}^{T-p} \hat{\underline{x}}_{it}^*\hat{\underline{y}}_{it}^*\right] = \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{x}}_t\mathbf{\Omega}_{uu}^{-1}\hat{\mathbf{x}}_t'\right]^{-1} \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{x}}_t\mathbf{\Omega}_{uu}^{-1}\hat{\underline{y}}_t\right].$$

The properties of this estimator are given in the following corollary to proposition 1.[4]

Corollary 1 (Asymptotic distribution of restricted DSUR). Let $\mathbf{b}_e = (\underline{B}_{e_1}, \underline{B}_{e_2})$, $\mathbf{R}$ be a $q \times 2k$ matrix of constants such that $\mathbf{R}\hat{\underline{\beta}}_{rdsur} = \underline{r}$, and $\hat{\mathbf{V}}_{rdsur} = \sum_{t=p+1}^{T-p} \mathbf{x}_t\mathbf{\Omega}_{uu}^{-1}\mathbf{x}_t'$. Then as $T_* \to \infty$,

$$T_*(\hat{\underline{\beta}}_{rdsur} - \underline{\beta}) \xrightarrow{D} \left(\int \mathbf{b}_e\mathbf{\Omega}_{uu}^{-1}\mathbf{b}_e'\right)^{-1} \left(\int \mathbf{b}_e\mathbf{\Omega}_{uu}^{-1}d\underline{B}_u\right), \tag{16}$$

and

$$(\mathbf{R}\hat{\underline{\beta}}_{rdsur} - \underline{r})' \left[\mathbf{R}\hat{\mathbf{V}}_{rdsur}\mathbf{R}'\right]^{-1} (\mathbf{R}\hat{\underline{\beta}}_{rdsur} - \underline{r}) \xrightarrow{D} \chi_q^2. \tag{17}$$

[4] In matrix notation, let $\hat{\underline{Y}}_{\mathsf{T}} = (\hat{\underline{Y}}_1', \hat{\underline{Y}}_2')'$ where $\hat{\underline{Y}}_{\mathsf{i}} = (\hat{\underline{y}}_{\mathsf{ip+1}}, \ldots, \hat{\underline{y}}_{\mathsf{iT-p}})'$, $\hat{\mathbf{X}}_{\mathsf{T}} = (\hat{\mathbf{X}}_1', \hat{\mathbf{X}}_2')'$, $\hat{\mathbf{X}}_{\mathsf{i}} = (\hat{\underline{x}}_{\mathsf{ip+1}}, \ldots, \hat{\underline{x}}_{\mathsf{iT-p}})'$ is the $T_* \times k$ matrix of regressors, and $\hat{\underline{u}}_{\mathsf{T}} = (\hat{\underline{u}}_1, \hat{\underline{u}}_2)'$, $\hat{\underline{u}}_{\mathsf{i}} = (\hat{u}_{\mathsf{ip+1}}, \ldots \hat{u}_{\mathsf{iT-p}})'$. The stacked system of observations is $\hat{\underline{Y}}_{\mathsf{T}} = \hat{\mathbf{X}}_{\mathsf{T}}\underline{\beta} + \hat{\underline{u}}_{\mathsf{T}}$ where $\hat{\underline{\beta}}_{\mathsf{rdsur}} = [\hat{\mathbf{X}}_{\mathsf{T}}'(\mathbf{\Omega}_{\mathsf{uu}}^{-1} \otimes \mathbf{I}_{\mathsf{T}})\hat{\mathbf{X}}_{\mathsf{T}}]^{-1}[\hat{\mathbf{X}}_{\mathsf{T}}'(\mathbf{\Omega}_{\mathsf{uu}}^{-1} \otimes \mathbf{I}_{\mathsf{T}})\hat{\underline{Y}}_{\mathsf{T}}]$.

1.2.2 Panel DOLS

In panel DOLS, control for cross-equation endogeneity can also be achieved by working with first-step errors from regressing y_{it} and each element of $\underline{x}_{it}$ on $\underline{z}_{pt}$. Using 'hats' to denote the resulting least-squares residuals, the panel DOLS estimator is,

$$\hat{\underline{\beta}}_{pdols} = \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{x}}_t \hat{\mathbf{x}}_t'\right]^{-1} \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{x}}_t \hat{\underline{y}}_t\right], \tag{18}$$

where $\hat{\mathbf{x}}_t = (\hat{\underline{x}}_{1t}, \hat{\underline{x}}_{2t})$ is a $k \times 2$ matrix. The asymptotic sampling properties of panel DOLS under cross-sectional dependence are given as a corollary to proposition 3.

Corollary 2 (Asymptotic distribution of panel DOLS). Let $\mathbf{b}_e = (\underline{B}_{e_1}, \underline{B}_{e_2})$, $\hat{\mathbf{V}}_{pdols} = \left[\sum_{t=p+1}^{T-p} \mathbf{x}_t \mathbf{x}_t'\right]^{-1} \left[\sum_{t=p+1}^{T-p} \mathbf{x}_t \mathbf{\Omega}_{uu} \mathbf{x}_t'\right] \left[\sum_{t=p+1}^{T-p} \mathbf{x}_t \mathbf{x}_t'\right]^{-1}$, and $\mathbf{R}$ be a $q \times 2k$ matrix of constants such that $\mathbf{R}\underline{\beta} = \underline{r}$. Then as $T_* \to \infty$,

$$T_*(\hat{\underline{\beta}}_{pdols} - \underline{\beta}) \xrightarrow{D} \left(\int \mathbf{b}_e \mathbf{b}_e'\right)^{-1} \int \mathbf{b}_e d\underline{B}_u, \tag{19}$$

and

$$(\mathbf{R}\hat{\underline{\beta}}_{pdols} - \underline{r})' \left[\mathbf{R}\hat{\mathbf{V}}_{pdols}\mathbf{R}'\right]^{-1} (\mathbf{R}\hat{\underline{\beta}}_{pdols} - \underline{r}) \xrightarrow{D} \chi_q^2. \tag{20}$$

Finally, it should be obvious that avar$(\hat{\underline{\beta}}_{rdsur}) \leq$ avar$(\hat{\underline{\beta}}_{pdols})$.

2 Monte Carlo Experiments

In this section, we study the small sample properties of the two-step estimators discussed above by way of a series of Monte Carlo experiments. Section 2.1 describes the data generating process and the estimation procedures that we use. Section 2.2 reports the results. First, we compare the performance of DSUR, feasible DSUR, system and ordinary DOLS methods in an environment where the cointegration vector exhibits heterogeneity across equations. Second, we compare restricted DSUR, feasible restricted DSUR, and panel DOLS in an environment where the cointegrating vector is identical across equations.

2.1 Experimental Design

The cointegrating regression has a single regressor. The general form of the data generating process (DGP) is given by,

$$\begin{aligned} y_{it} &= x_{it}\beta_i + u_{it}^{\dagger}, \quad i = 1, 2, & (21) \\ \Delta x_{it} &= e_{it}, & (22) \\ \underline{\eta}_t &= \mathbf{A}\underline{\eta}_{t-1} + \underline{\epsilon}_t, & (23) \end{aligned}$$

where $\underline{\eta}_t = (u_{1t}^{\dagger}, u_{2t}^{\dagger}, e_{1t}, e_{2t})'$, $\underline{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t}, \epsilon_{4t})' \overset{iid}{\sim} N(0, \mathbf{\Sigma})$ and $\mathbf{A}$ is a 4×4 matrix of coefficients. Observations are generated under alternative specifications that differ by the degree of cross-sectional dependence and by the innovation variances of the equilibrium errors. We consider the following six cases.

Case I builds in 'own equation' endogeneity but no cross-sectional endogeneity. That is, $u_{it}^{\dagger}$ is correlated with leads and lags of e_{jt} for $i = j$ but not for $i \neq j$. We allow only contemporaneous cross-sectional dependence in the equilibrium errors $u_{1t}^{\dagger}$ and $u_{2t}^{\dagger}$. This is achieved by setting

$$\mathbf{A}_1 = \begin{bmatrix} 0.90 & 0.0 & 0.05 & 0.0 \\ 0.0 & 0.90 & 0.0 & 0.05 \\ 0.05 & 0.0 & 0.25 & 0.0 \\ 0.00 & 0.05 & 0.0 & 0.25 \end{bmatrix}, \quad \mathbf{\Sigma}_1 = \begin{bmatrix} 1 & 0.2 & 0 & 0 \\ 0.2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Case II introduces 'cross-equation' endogeneity by making $u_{it}^{\dagger}$ correlated with leads and lags of $e_{jt}, (i, j = 1, 2)$ by setting

$$\mathbf{A}_2 = \begin{bmatrix} 0.90 & 0.0 & 0.05 & -0.05 \\ 0.0 & 0.90 & -0.05 & 0.05 \\ 0.05 & -0.05 & 0.25 & 0.0 \\ -0.05 & 0.05 & 0.0 & 0.25 \end{bmatrix}, \quad \mathbf{\Sigma}_2 = \mathbf{\Sigma}_1.$$

Case III intensifies the degree of contemporaneous cross-equation correlation of the equilibrium errors by setting

$$\mathbf{A}_3 = \mathbf{A}_2, \quad \mathbf{\Sigma}_3 = \begin{bmatrix} 1 & 0.8 & 0 & 0 \\ 0.8 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The next three cases introduce differences between the innovation variances for the equilibrium errors. Cases IV, V, and VI are identical to cases I, II, and III respectively except the innovation variance of $u_{1t}^{\dagger}$ is 10 times larger than the innovation variance of $u_{2t}^{\dagger}$. The original correlation between the innovations is preserved. Specifically,

Case IV.

$$\mathbf{A}_4 = \mathbf{A}_1, \quad \mathbf{\Sigma}_4 = \begin{bmatrix} 10 & 0.632 & 0 & 0 \\ 0.632 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Case V. $\mathbf{A}_5 = \mathbf{A}_2$, and $\mathbf{\Sigma}_5 = \mathbf{\Sigma}_4$.

Case VI.

$$\mathbf{A}_6 = \mathbf{A}_2, \quad \mathbf{\Sigma}_6 = \begin{bmatrix} 10 & 2.53 & 0 & 0 \\ 2.53 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

For each experiment, we generate 10,000 random samples of T observations. Under heterogeneous cointegration, $\beta_1 = 1.4$ and $\beta_2 = 0.6$. Under homogeneous cointegration, we set $\beta_1 = \beta_2 = \beta = 1.0$. To purge the effects of endogeneity in the system estimators, first-step regressions are run including p leads and lags of Δx_{1t} and Δx_{2t} in each equation. For ordinary DOLS, we include p leads and lags only of the 'own' Δx_{it}.

An important problem in applications is how to choose p. Unfortunately, no standard method has emerged even for time series. Often, the ad hoc rule used by Stock and Watson (1993) that sets $p = 1$ for $T = 50$, $p = 2$ for $T = 100$, and $p = 3$ for $T = 300$ is adopted in Monte Carlo and empirical studies. While it is desirable to have a data dependent method, such as an information criterion or general-to-specific rules for choosing p, such rules quickly become unwieldy as the size of the cross-section grows. To balance concerns for employing a data dependent method in applications, evaluation of estimator performance, and manageability of the method, we apply the following modified BIC rule to choose p: Let $p_{ij}^{+}(p_{ij}^{-})$ denote the number of leads (lags) of $\Delta \underline{x}_j$ in equation i. First run DOLS and determine (p_{ii}^{+}, p_{ii}^{-}) by minimizing BIC, then for $i \neq j$, set $(p_{ij}^{+}, p_{ij}^{-}) = (p_{ii}^{+}, p_{ii}^{-})$.

The DSUR estimators are computed using the known long-run covariance matrix $\mathbf{\Omega}_{uu}$. Feasible DSUR is computed with a parametrically estimated $\mathbf{\Omega}_{uu}$. To do this, we model the residuals from first-step regressions as a restricted vector autoregression in which the individual residual processes are m-th ordered autoregressions. While an unrestricted vector autoregression might seem to be a more appropriate choice and is feasible in our two-equation example DGP, it quickly becomes too heavily parameterized in even moderately sized systems. Since the restricted VAR is a popular method for achieving model parsimony, we adopt that approach here.[5] Thus, let $M = \max(m_1, m_2)$, where m_i is the order of the autoregression for u_{it}, which we determine by the general-to-specific t-test method suggested by Hall (1994). For $t = 1, \ldots, T - M$, the restricted VAR is, $\underline{u}_t = \sum_{j=1}^{M} \mathbf{\Phi}_j \underline{u}_{t-j} + \underline{\nu}_t$, where $\underline{u}_t = (u_{1t}, u_{2t})'$, $\underline{\nu}_t = (\nu_{1t}, \nu_{2t})'$, $\mathrm{E}(\underline{\nu}_t \underline{\nu}_t') = \mathbf{W}$, and $\mathbf{\Phi}_j$ is a (2×2) matrix of coefficients with zeros in the off-diagonal elements.[6] The autoregressions are then jointly estimated by iterated SUR and the estimated long-run covariance matrix is, $\hat{\mathbf{\Omega}}_{uu} = [\mathbf{I}_2 - \sum_{j=1}^{m} \hat{\mathbf{\Phi}}_j]^{-1} \hat{\mathbf{W}} [\mathbf{I}_2 - \sum_{j=1}^{m} \hat{\mathbf{\Phi}}_j']^{-1}$.

2.2 Results

Table 1 reports 5, 50, and 95 percentiles and the mean of the Monte Carlo distribution for the estimators along with the relative (to DOLS) mean-square error. In case I where there is no cross-sectional endogeneity and a low degree of cross-sectional correlation, there is little difference among the estimators. None exhibit substantial bias and for $T = 100, 300$, are similar in terms of efficiency. The loss of efficiency involved in estimating the long-run covariance matrix to do feasible DSUR is modest. For example, with $T = 100$, the relative mean-square error for feasible DSUR is 1.04. At $T = 300$, we begin to see evidence of DSUR efficiency gains with relative mean-square error of 0.99. DSUR performance under case II, where cross-equation endogeneity is introduced, is slightly improved in terms of mean square error.

We observe substantial efficiency gains to using DSUR in case III, where there is a high degree of cross-equation correlation. For $T = 50$, DSUR achieves a 54 percent reduction in mean-square error over the system DOLS estimator. Similarly, feasible DSUR achieves a 31 percent reduction in mean-square error. These efficiency gains grow when $T = 300$. All of the estimators exhibit some upward bias in small samples. The bias is slightly more severe for DSUR. There is little difference in bias between DSUR and feasible DSUR.

We conclude from Table 1 that substantial efficiency gains can be achieved with DSUR over DOLS when there is a high degree of cross-equation dependence in the equilibrium errors. The results for cases IV-VI are nearly identical and are not reported

[5] This estimator of $\mathbf{\Omega}_{uu}$ is consistent if $M \to \infty$ as $T \to \infty$ and $M = o(T)$. This is true even if the zero-restrictions on the off-diagonal elements of $\mathbf{\Phi}_j$ are false [e.g., Andrews and Monahan (1992)].

[6] The general-to-specific method proceeds as follows: Start with some maximal lag order ℓ and estimate the autoregression on $\hat{u}_{it}$. Let $\hat{\phi}_{i\ell}$ be the $ii-$th element of $\mathbf{\Phi}_\ell$. If the absolute value of the t-ratio for $\hat{\phi}_{i\ell}$ is less than some appropriate critical value, c^*, reset m_i to $\ell - 1$ and reestimate. Repeat the process until the t-ratio of the estimated coefficient with the longest lag exceeds the critical value c^*.

to save space.

We now turn to the small-sample properties of Wald test statistics for the test of homogeneity, $\beta_1 = \beta_2$ in the cointegrating regression slope coefficient. Table 2 displays the 90, 95, and 99 percentiles of the test statistic and the percentile of the Monte Carlo distribution that lies to the right of the asymptotic distribution's 5% critical value (indicated by size (5%)) for system DOLS and DSUR. It can be seen that the DSUR test is uniformly and substantially more accurately sized than the system DOLS test. Moreover, the performance of the DSUR test and its relation to the system DOLS test is largely invariant to changes in the strength of the cross-sectional dependence or the relative size of the equilibrium error innovation variances.

Next, we consider test statistic performance in tests of the null hypothesis $H_0 : \beta_1 = \beta_2 = 1$. Table 3 reports the results for this experiment. Again, it can be seen that the DSUR test has better small-sample size properties than the system DOLS test.

We now consider estimation under homogeneity of the cointegration vector across equations. The small-sample performance of the restricted panel estimators, panel DOLS and restricted DSUR is reported in table 4. There is little difference in estimator performance in cases I and II while restricted DSUR and feasible restricted DSUR achieve substantial efficiency gains over panel DOLS in all other cases. The efficiency gain in restricted DSUR is more dramatic when there are differences in the innovation variance of the equilibrium errors across equations. In case VI for example, for $T = 50$, the mean square error of the restricted DSUR distribution is 73 percent lower than that of the panel DOLS distribution and the mean-square error the feasible restricted DSUR distribution lies 59 percent below that of panel DOLS. The rather large gaps in efficiency between restricted DSUR and panel dynamic OLS remain present even when $T = 300$.

We conclude that for $T = 300$, substantial efficiency gains are available for the DSUR methods, especially when there is moderate to strong cross-sectional dependence. For $T = 50, 100$, the tests of homogeneity restrictions are somewhat oversized and use of the asymptotic theory in applications may lead to over-rejections of the null hypothesis. With $T = 300$, the DSUR tests are reasonably sized.

3 Applications

In this section we illustrate the usefulness of DSUR by applying it to two empirical problems in international economics. Our first application revisits the anomaly reported by Evans and Lewis (1993) that the expected excess return from forward foreign exchange rate speculation is unit-root nonstationary. Our second application revisits the Feldstein and Horioka (1980) problem of estimating the correlation between national saving rates and national investment rates and the interpretation of this correlation as a measure of international capital mobility.

3.1 Spot and Forward Exchange Rates

Let s_{it} be the logarithm of the spot exchange rate between the home country and country i, and let f_{it} be the associated 1-period forward exchange rate. It is widely agreed that since the move to generalized floating in 1973 that both $s_{it} \sim I(1)$ and $f_{it} \sim I(1)$ and that they are cointegrated. Let β_i be the cointegrating coefficient between s_{it+1} and f_{it} and let $p_{it} = f_{it} - \mathrm{E}_t(s_{it+1})$ be the expected excess return from forward foreign exchange speculation. The spot rate can be decomposed as $s_{it+1} = f_{it} - p_{it} + \epsilon_{it+1}$ where $\epsilon_{it+1} = s_{it+1} - \mathrm{E}_t(s_{it+1})$ is a rational expectations error, and the equilibrium error can be decomposed as $s_{it+1} - \beta_i f_{it} = (1-\beta_i)f_{it} - p_{it} + \epsilon_{it+1}$. If $\beta_i \neq 1$, it follows that the expected excess return p_{it} is nonstationary and is cointegrated with f_{it}. Evans and Lewis ask whether p_{it} is I(0) or I(1), by estimating the regression

$$s_{it+1} = \alpha_i + \beta_i f_{it} + u^{\dagger}_{it+1}, \tag{24}$$

by ordinary DOLS and testing the hypothesis $H_o : \beta_i = 1$. They use monthly observations from January 1975 through December 1989 on the dollar rates of the pound, deutschemark, and yen, are able to reject that the slope coefficient is 1 at small significance levels. The implied nonstationarity of the excess return is an anomaly.

We revisit the Evans and Lewis problem using an updated data set. Our data are spot and 30-day forward exchange rates for the pound, deutschemark, and yen relative to the U.S. dollar from January 1975 to December 1996. We obtain 286 time-series observations sampled from every 4th Friday of the Bank of Montreal/Harris Bank *Foreign Exchange Weekly Review*. Because all of the currency prices are in terms of a common numeraire currency, cross-equation error correlation is likely to be important. Under this setting, the regression errors are forecast errors of investors and will be correlated as long as information sets of investors in different countries contain common components.

The estimation results are reported in table 5. In light of the moderate size distortion uncovered in the Monte Carlo analysis, we test hypotheses using the 1 percent asymptotic significance level. Our BIC rule recommends including $p = 3$ leads and lags of the endogeneity control variables. The DSUR estimates with $p = 3$ are insignificantly different from 1 for the pound and yen, but is significantly less than 1 for the deutschemark. We employ two tests of homogeneity in the cointegration vectors. The first one tests the null hypothesis $H_o : \beta_1 = \beta_3, \beta_2 = \beta_3$. The second is a test of the null hypothesis $H_o : \beta_1 = \beta_2 = \beta_3 = 1$. These homogeneity restrictions cannot be rejected at the 1 percent level ($\chi^2_2 = 7.5$, p-value=0.024, $\chi^2_3 = 7.6$, p-value=0.056). We therefore proceed to impose the homogeneity restrictions in estimation and obtain a restricted

DSUR estimate that is insignificantly different from 1.

To investigate the sensitivity of the results to variations in the lead-lag specification used to control for endogeneity, we perform estimation with 2 leads and lags, and with 3 leads, and with 2 leads (no contemporaneous nor lagged values). The rationale for omitting the contemporaneous and lagged values of Δf_t is that under rational expectations if the forward exchange rate is the optimal predictor of the future spot rate, the equilibrium error $u^{\dagger}_{it+1}$ is orthogonal to any date t information. As can be seen, the results are qualitatively similar across the alternative lead-lag specifications. Here, as in many rational expectations models, it is more important to include leads than lags.

We conclude that the evidence for nonstationarity of the excess return is less compelling according to the DSUR slope coefficient estimates under homogeneity restrictions.

3.2 National Saving and Investment Correlations

Let $(I/Y)_i$ be the time-series average of the investment to GDP ratio in country i, and $(S/Y)_i$ be the analogous time-series average of the saving ratio to GDP ratio. Feldstein and Horioka (1980) run the cross-sectional regression,

$$\left(\frac{I}{Y}\right)_i = \alpha + \beta\left(\frac{S}{Y}\right)_i + u_i, \tag{25}$$

to test the hypothesis that capital is perfectly mobile internationally. They find that β is significantly greater than 0, and conclude that capital is internationally immobile.

The logic behind the Feldstein and Horioka regression goes as follows. Suppose that capital is freely mobile internationally. National investment should depend primarily on country-specific shocks. If the marginal product of capital in country i is high, it will attract investment. National saving on the other hand will follow investment opportunities not just at home, but around the world and will tend to flow towards projects that offer the highest (risk adjusted) rate of return. The saving rate in country i then is determined not by country–i specific events but by investment opportunities around the world. Under perfect capital mobility, the correlation between national investment and national saving should be low. Following the publication of Feldstein and Horioka's cross-sectional study, a number of follow-up cross-sectional and panel studies have reported that national saving rates are highly correlated with national investment rates [For surveys of the Feldstein–Horioka literature, see Bayoumi (1997) and Coakley et al. (1998)].

Theoretical studies, on the other hand, have shown that The Feldstein–Horioka (1980) logic is not airtight. Obstfeld (1986), Cantor and Mark (1988), Cole and Obstfeld (1991) Baxter and Crucini (1993) provide counterexamples in which the economic environment is characterized by perfect capital mobility but decisions by optimizing agents lead to highly correlated saving and investment rates. Along with theoretical criticism against the Feldstein and Horioka hypothesis, more than a dozen empirical studies have criticized their econometrics by arguing that the saving and investment ratios are non-stationary.

Coakley *et al.* (1996) suggest an alternative interpretation of the long-run relationship between saving and investment. By the national income accounting identity, the difference between national investment and national saving is the current account balance. Coakley *et al.* argue that the current account must be stationary when the present value of expected future debt acquisition is bounded. In other words, whether the current account balance is stationary depends not on the degree of capital mobility but on whether the long-run solvency constraint holds. If saving and investment are unit root nonstationary, they are cointegrated with a cointegrating vector (1,-1). Thus the long-run relationship between saving and investment studied by time series cointegrating regressions is best interpreted as a test of the long-run solvency constraint and not of the degree of capital mobility. Furthermore, Coakley and Kulasi (1997), Hussein (1998), and Jansen (1996) show that the saving and investment ratios are cointegrated.

We employ DSUR to re-examine the Feldstein–Horioka puzzle using 100 quarterly observations from the *International Financial Statistics* CD-ROM on nominal GDP, saving, and investment from 1970.1 to 1995.4 for Australia, Austria, Canada, Finland, France, Germany, Italy, Japan, Spain, Switzerland, the U.K., and the U.S. Since our focus is on the long-run relationship between saving and investment, we follow Coakley *et al.*'s interpretation that the long-run solvency constraint implies cointegration. Even though Coakley *et al.* do not emphasize this, we note that two versions of their model imply slightly different forms of cointegration. First, if we assume that saving and investment are unit root nonstationary, then this version of their model implies that the current account is stationary and saving and investment are cointegrated with a cointegrating vector of (1,-1). Second, if we assume that saving-GDP ratio and investment-GDP ratio are unit root nonstationary, we must interpret saving and investment in their model to be normalized by GDP. The second version of their model implies that the current account over GDP is stationary and that saving and investment normalized by GDP are cointegrated with a cointegrating vector of (1,-1).

For the first version of the model, we run the regression in levels after normalizing saving and investment by GDP,

$$\left(\frac{I}{Y}\right)_{it} = \alpha_i + \beta_i \left(\frac{S}{Y}\right)_{it} + u_{it}. \tag{26}$$

Presumably, the reason for normalizing investment and saving by GDP in many applications is to transform the data into stationary observations, as they would be if the economy is on a balanced growth path. However, we find very little empirical evidence for this implication of the balanced growth in our data set.[7]

For the second version of Coakley et al.'s model, we run the regression should in log levels,

$$\ln(I_{it}) = \alpha_i + \beta_i \ln(S_{it}) + u_{it}. \tag{27}$$

In both versions, the cross-equation error correlation is likely to be important because the error for each country is an infinite sum of shocks to saving and investment. There is an additional reason for the correlation to be important in the second version of the model because normalizing by GDP can create artificial correlation between the ratios even when the levels are uncorrelated. An income shock automatically affects both (I/Y) and (S/Y) independently of its effect on investment and saving thus generating artificial correlation between the ratios.

It was not feasible for us to simultaneously estimate the regressions for all 12 countries due to the excessive number of parameters that needed to be estimated to implement DSUR. To proceed, we break the panel into subsamples and estimate separate systems for European and non European countries.

Table 6 reports our estimates of the regression. We look first at the results in ratio form. For the European countries, the BIC rule selects $p = 3$. We obtain DSUR slope coefficients estimates that lie below 1 for the UK, Spain and Germany, estimates that are near 1 for France and Austria, and estimates that significantly exceed 1 for Finland, Italy, and Switzerland. For non European countries ($p = 3$), the point estimates are insignificantly different for 1 for the U.S., Canada, and Japan. Only the estimate for Australia is significantly less than 1.

Tests of homogeneity are mixed. In the European system, the asymptotic p-values for the test of homogeneity and also for the test that all slope coefficients are 1 are both 0.000. For the non-European system, neither of the tests for homogeneity can be rejected at the asymptotic 1 percent level. These results suggest that for the non-European system, it is reasonable to pool and to re-estimate under homogeneity. When we do so, we obtain a restricted DSUR estimate 0.78 which is significantly less than 1.

Looking at the estimates from the log-levels regression, the European data set tells a mixed story. These estimates are associated with $p = 3$. The point estimates for

[7] We perform Phillips and Sul's (2002) panel unit root test which are robust to cross-sectional dependence. Their suggestion is to apply an orthogonalization procedure to the observations under the assumption that the cross-sectional dependence is generated by a factor structure, and then to apply the Maddala–Wu (1999) panel unit-root test to the orthogonalized observations. The series tested and associated p-values from the tests are as follows: $S/Y, (0.972)$, $I/Y, (0.999)$, $\ln(S), (1.000)$, $\ln(I), (1.000)$. Since none of the p-values are less than 0.05, the null hypothesis of a unit root is not rejected. In differences, we obtain for $(S-I)/Y, (0.000)$, and $\ln(S/I)(0.000)$ and are able to reject the unit root null hypothesis for these cases.

Switzerland and Finland are significantly less than 1, but the Wald test does not reject the homogeneity restriction at any reasonable level. As a result, we pool and re-estimate under homogeneity restrictions on the slope coefficient with restricted DSUR and obtain a point estimate of 0.97, which is insignificantly different from 1. In the log-levels regression for the non-European countries, our BIC rule sets $p = 2$. Here, only the DSUR estimate for the US of 1.10 is significantly greater than 1. The homogeneity restrictions are not rejected so we pool and obtain a restricted DSUR estimate of 1.02 which is insignificant different from 1.

To summarize, the weight of the evidence suggests that the long-run slope coefficients in the saving–investment regressions are very close to 1 for most countries which is consistent with the hypothesis that Coakley et al.'s solvency constraint is not violated.

4 Conclusion

In this paper, we proposed the dynamic seemingly unrelated regression estimator for multiple-equation cointegrating regressions both in situations when the cointegration vector displays heterogeneity across equations and when it is homogeneous. This estimator exploits the cross-equation correlation in the errors, is asymptotically efficient, and is computationally more convenient to use than the existing nonparametric versions of seemingly unrelated cointegrating regression estimators. Our Monte Carlo studies suggest that the small sample properties conform largely according to the predictions of the asymptotic theory. In most of the cases that we examined, DSUR estimators are more efficient than DOLS estimators which do not utilize the cross-equation correlation. The efficiency gain is increasing in the correlation of the equilibrium errors across equations. In the case of homogenous cointegrating vectors, the efficiency gain is also increasing in the difference between in the error variance across equations. These results stand in contrast to Park and Ogaki's (1991) seemingly unrelated CCR estimators, which also are asymptotically efficient, but in small samples were found in many cases to be less efficient than equation-by-equation CCR estimators.

We showed that these estimators can be successfully applied in small to moderate systems where the number of time periods, T, is substantially larger than the number of equations, N. DSUR will not be computationally feasible in systems of large N because the number of free parameters that must be estimated in the error correlation quickly becomes unwieldy as N grows. In the foreign exchange rate application, N is 3 and this size condition is satisfied. However, in the saving-investment regression, we found it necessary to split up the sample. We did so according to geography so that each subsample might reasonably exhibit different levels of cross-equation error correlation.

Finally, we have stressed the computational convenience of DSUR for correcting endogeneity in small nonstationary panels as an advantage over nonparametric methods such as those suggested in Park and Ogaki (1991) and Moon (1999). The alternative approaches involve an age-old tradeoff to the researcher. The lack of computational transparency of the nonparametric methods may be viewed as the price of flexibility whereas the computational tractability of the parametric method creates the possibility for misspecification error, which we did not explicitly consider in the paper.

References

Andrews, Donald W.K. and J. Christopher Monahan (1992). 'An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator,' Econometrica, 60(4) July, pp. 953-966.

Baxter, Marianne and Mario Crucini (1993). 'Explaining Saving-Investment Correlations,' American Economic Review, 83(3) June, pp. 416–36.

Bayoumi, Tamim. (1997) Financial Integration and Real Activity, Ann Arbor, University of Michigan Press.

Cantor, Richard and Nelson C. Mark (1988). 'The International Transmission of Real Business Cycles," International Economic Review, 29,3 August. pp. 493-507.

Coakley, Jerry and Farida Kulasi (1997). 'Cointegration of Long Run Saving and Investment,' Economics Letters, 54 pp 1–6.

Coakley, Jerry, Farida Kulasi and Ron Smith (1998). 'The Feldstein-Horioka Puzzle and Capital Mobility: A Review,' International Journal of Finance and Economics 3, pp. 169–188.

Coakley, Jerry, Farida Kulasi and Ron Smith (1996). 'Current Account Solvency and the Feldstein-Horioka Puzzle,' Economic Journal 106 (May), pp. 620–627.

Cole, Harold and Maurice Obstfeld. (1991). 'Commodity Trade and International Risk Sharing: How Much Do Financial Markets Matter?' Journal of Monetary Economics 28(1), August, pp. 3–24.

Dekle, Robert (1996). 'Saving–Investment Associations and Capital Mobility: On the Evidence from Japanese Regional Data.' Journal of International Economics 41, pp.53–72.

Evans, Martin D.D. and Karen K. Lewis (1995). 'Do Long-Term Swings in the Dollar Affect Estimates of the Risk Premia?' Review of Financial Studies, Fall, 8(3), pp. 709–742.

Feldstein, Martin and Charles Horioka (1980). 'Domestic Saving and International Capital Flows,' Economic Journal, 90(2), pp. 314–29.

Hall, Alistar (1994). 'Testing for a Unit Root in Time Series with Pretest Data-Based Model' Journal of Business and Economic Statistics 12: 461-70.

Hussein, Khaled A. (1998). 'International Capital Mobility in OECD Countries: The Feldstein–Horioka 'Puzzle' Revisited,' Economics Letters 59, pp. 237–242.

Jansen, Jos W. (1996). 'Estimating Saving-investment Correlations: Evidence for OECD Countries Based on an Error Correction Model,' Journal of International Money and Finance, 15(5), pp. 749–781.

Im, Kyung So, M. Hashem Pesaran and Yongcheol Shin. 'Testing For Unit Roots In Heterogeneous May,' Panels 1995, Revised March 2002. Journal of Econometrics (forthcoming)

Kyung So Im, M. Hashem Pesaran and Yongcheol, (1997) Shin. 'Testing for Unit Roots in Heterogeneous Panels,' mimeo Trinity College, University of Cambridge.

Kao, Chihwa and Min-Hsien Chiang (1998). 'On the Estimation and Inference of a Cointegrated Regression in Panel Data,' mimeo, Syracuse University.

Kao, Chihwa (1999). 'Spurious Regression and Residual-Based Tests for Cointegration in Panel Data,' Journal of Econometrics, 90, pp. 1-44.

Maddala, G.S. and Shaowen Wu (1999). "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test," Oxford Bulletin of Economics and Statistics, Special Issue 0305-9049, pp. 631–52.

Mark, Nelson C. and Donggyu Sul, (1999). 'A Computationally Simple Cointegration Vector Estimator for Panel Data,' mimeo, Ohio State University.

Moon, R. Hyungsik (1999). 'A Note on Fully-Modified Estimation of Seemingly Unrelated Regressions Models with Integrated Regressors,' (Economics Letters), 65, 25-31.

Moon, R. Hyungsik and Benoit Perron (2000). 'The Seemingly Unrelated Dynamic Cointegration Regression Model and Testing for Purchasing Power Parity,' mimeo, University of Southern California.

Obstfeld, Maurice (1986). 'Capital Mobility in the World Economy: Theory and Measurement,' Carnegie–Rochester Conference Series on Public Policy, Spring 1986, 24, pp. 55–104.

Park, Joon Y. (1992). 'Canonical Cointegrating Regressions,' Econometrica 60(1), pp. 119-43.

Park, Joon Y. and Masao Ogaki (1991). 'Seemingly Unrelated Canonical Cointegrating Regressions,' Rochester Center for Economic Research Working Paper No. 280.

Park, Joon and Peter C.B. Phillips (1988). 'Statistical Inference in Regressions with Integrated Processes: Part 1,' Econometric Theory, 4, pp 468–497.

Pedroni, Peter (1997). 'Fully Modified OLS for Heterogeneous Cointegrated Panels and the Case of Purchasing Power Parity,' mimeo, Department of Economics, Indiana University.

Phillips, Peter C.B. and Bruce E. Hansen. (1990). 'Statistical Inference in Instrumental Variables Regression with I(1) Processes.' Review of Economic Studies 57, pp. 99–125.

Phillips, Peter C.B. and Mico Loretan. (1991). 'Estimating Long-Run Economic Equilibria.' Review of Economic Studies, 58, pp.407–36.

Phillips, Peter C. B. and Hyungsik R. Moon (1998). 'Linear Regression Limit Theory for Nonstationary Panel Data,' Econometrica forthcoming.

Phillips, Peter C.B. and Donggyu Sul (2002). "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence," mimeo, Yale University.

Saikkonen, Pentti (1991). 'Asymptotically Efficient Estimation of Cointegration Regressions,' Econometric Theory 7:1–21.

Sims, Christopher A., James H. Stock, and Mark W. Watson (1990). 'Inference in Linear Time Series Models with Some Unit Roots,' Econometrica 58: 113-44.

Stock, James H. and Mark W. Watson (1993). 'A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems,' Econometrica 61:783–820.

Table 1: Monte Carlo Performance of DOLS and DSUR Estimators under Cointegration Vector Heterogeneity, Cases I-III.

		$\beta_1 = 1.4$					$\beta_2 = 0.6$				
	T	5%	50%	95%	mean	Rel. MSE	5%	50%	95%	mean	Rel. MSE
Case I.											
DOLS	50	0.818	1.405	1.973	1.401	1.000	0.031	0.603	1.192	0.606	1.000
SDOLS	50	0.768	1.404	2.020	1.400	1.152	-0.010	0.604	1.242	0.607	1.196
DSUR	50	0.772	1.405	2.011	1.401	1.127	-0.004	0.603	1.239	0.608	1.164
FDSUR	50	0.756	1.407	2.033	1.402	1.223	-0.026	0.602	1.254	0.608	1.255
DOLS	100	1.103	1.400	1.701	1.401	1.000	0.294	0.599	0.899	0.600	1.000
SDOLS	100	1.096	1.398	1.707	1.401	1.025	0.291	0.599	0.910	0.600	1.042
DSUR	100	1.101	1.400	1.699	1.401	1.004	0.296	0.599	0.907	0.600	1.018
FDSUR	100	1.099	1.400	1.707	1.402	1.037	0.294	0.598	0.913	0.601	1.057
DOLS	300	1.296	1.399	1.504	1.400	1.000	0.497	0.600	0.703	0.600	1.000
SDOLS	300	1.295	1.400	1.503	1.399	1.004	0.497	0.600	0.704	0.600	1.001
DSUR	300	1.297	1.400	1.504	1.400	0.972	0.500	0.600	0.702	0.601	0.975
FDSUR	300	1.296	1.400	1.504	1.400	0.989	0.499	0.600	0.703	0.601	0.986
Case II.											
DOLS	50	0.824	1.407	1.955	1.401	1.000	0.005	0.607	1.196	0.606	1.000
SDOLS	50	0.788	1.408	1.998	1.402	1.171	-0.014	0.608	1.241	0.610	1.163
DSUR	50	0.806	1.407	1.984	1.403	1.115	-0.010	0.605	1.234	0.610	1.113
FDSUR	50	0.779	1.409	1.998	1.403	1.196	-0.024	0.607	1.251	0.609	1.215
DOLS	100	1.111	1.400	1.701	1.404	1.000	0.311	0.600	0.891	0.601	1.000
SDOLS	100	1.115	1.400	1.699	1.403	0.981	0.316	0.599	0.890	0.601	0.987
DSUR	100	1.124	1.402	1.686	1.404	0.931	0.323	0.599	0.883	0.602	0.937
FDSUR	100	1.117	1.401	1.697	1.403	0.968	0.317	0.600	0.887	0.602	0.975
DOLS	300	1.304	1.400	1.500	1.401	1.000	0.505	0.601	0.700	0.602	1.000
SDOLS	300	1.306	1.400	1.496	1.400	0.948	0.507	0.601	0.698	0.601	0.946
DSUR	300	1.309	1.400	1.494	1.401	0.881	0.511	0.601	0.695	0.602	0.878
FDSUR	300	1.307	1.400	1.495	1.400	0.908	0.509	0.601	0.696	0.602	0.905
Case III.											
DOLS	50	0.796	1.420	2.035	1.418	1.000	-0.015	0.619	1.235	0.620	1.000
SDOLS	50	0.753	1.420	2.067	1.417	1.159	-0.042	0.617	1.293	0.622	1.146
DSUR	50	0.987	1.430	1.883	1.430	0.536	0.185	0.630	1.083	0.631	0.539
FDSUR	50	0.889	1.427	1.969	1.426	0.796	0.080	0.627	1.173	0.628	0.824
DOLS	100	1.085	1.409	1.737	1.411	1.000	0.286	0.610	0.937	0.613	1.000
SDOLS	100	1.085	1.410	1.734	1.411	1.003	0.290	0.610	0.939	0.613	0.998
DSUR	100	1.210	1.419	1.641	1.421	0.451	0.406	0.618	0.846	0.620	0.454
FDSUR	100	1.185	1.418	1.659	1.421	0.564	0.384	0.617	0.866	0.620	0.569
DOLS	300	1.292	1.404	1.522	1.404	1.000	0.493	0.603	0.717	0.604	1.000
SDOLS	300	1.292	1.403	1.521	1.404	0.993	0.493	0.603	0.717	0.603	0.996
DSUR	300	1.337	1.407	1.485	1.408	0.439	0.536	0.607	0.685	0.608	0.455
FDSUR	300	1.333	1.407	1.488	1.408	0.472	0.533	0.607	0.687	0.608	0.491

Notes: SDOLS is system DOLS, FDSUR is feasible DSUR, Rel. MSE is relative (to DOLS) mean square error.

Table 2: Monte Carlo Performance of Tests of the Homogeneity Restriction $H_0 : \beta_1 = \beta_2$.

		SDOLS				DSUR			
Case	T	90%	95%	99%	size (5%)	90%	95%	99%	size (5%)
I	50	8.108	15.872	56.751	0.545	13.524	22.221	51.059	0.323
	100	4.820	8.491	25.572	0.299	5.789	8.793	19.928	0.174
	300	3.165	5.562	12.688	0.135	3.587	5.281	10.274	0.089
II	50	10.451	19.752	66.978	0.557	14.661	23.066	54.308	0.340
	100	6.721	12.005	33.957	0.310	6.423	10.179	22.492	0.189
	300	5.019	8.230	19.430	0.144	4.101	6.112	11.206	0.109
III	50	10.079	18.049	53.630	0.512	12.592	19.587	47.843	0.305
	100	5.750	9.702	26.575	0.266	5.621	8.623	18.579	0.159
	300	4.507	6.955	14.345	0.115	3.558	5.242	9.768	0.091
IV	50	7.450	15.209	58.422	0.529	13.135	20.853	51.627	0.312
	100	4.583	8.880	24.810	0.289	5.704	8.605	18.208	0.169
	300	3.098	5.626	14.635	0.132	3.708	5.335	10.065	0.094
V	50	12.014	23.684	89.429	0.538	14.799	22.897	50.549	0.338
	100	8.672	16.130	50.511	0.305	7.165	11.200	23.682	0.207
	300	7.380	12.264	28.027	0.143	4.888	7.119	14.041	0.141
VI	50	12.258	25.374	96.759	0.515	12.572	19.806	43.528	0.313
	100	8.183	15.254	47.253	0.286	5.735	8.529	17.390	0.165
	300	6.234	11.472	28.683	0.148	3.717	5.448	10.060	0.095

Table 3: Monte Carlo Performance of DOLS and DSUR Tests of the Homogeneity Restriction $H_0 : \beta_1 = \beta_2 = 1$

		DOLS				DSUR			
Case	T	90%	95%	99%	size (5%)	90%	95%	99%	size (5%)
I	50	58.249	97.720	244.903	0.541	39.327	60.985	136.225	0.175
	100	19.569	31.885	70.288	0.338	13.819	20.517	45.418	0.125
	300	9.421	13.780	27.021	0.186	7.145	9.796	16.616	0.079
II	50	66.888	112.760	282.609	0.569	39.360	61.234	133.531	0.219
	100	24.383	38.798	79.687	0.374	14.517	21.949	41.870	0.168
	300	11.890	18.047	35.749	0.236	7.293	9.936	17.006	0.136
III	50	53.340	90.869	224.586	0.504	34.805	55.553	130.188	0.214
	100	16.649	26.686	60.757	0.296	12.612	18.255	36.911	0.152
	300	8.494	12.471	24.680	0.170	6.475	8.759	15.178	0.121
IV	50	98.368	168.956	499.771	0.584	34.962	55.504	122.012	0.159
	100	27.614	45.565	119.298	0.395	13.333	19.040	39.158	0.119
	300	11.986	18.400	38.804	0.235	7.037	9.626	16.079	0.080
V	50	129.155	223.862	581.759	0.630	36.696	57.173	121.274	0.224
	100	45.633	76.659	157.561	0.490	14.069	21.464	41.884	0.195
	300	22.422	33.981	70.126	0.366	7.434	10.280	19.574	0.184
VI	50	142.474	265.100	759.475	0.613	33.616	53.286	118.711	0.233
	100	48.326	84.267	207.843	0.459	13.286	19.014	38.011	0.190
	300	24.039	38.889	90.725	0.355	7.425	10.503	17.840	0.163

Table 4: Monte Carlo Performance of PDOLS and RDSUR Estimators under Cointegration Vector Homogeneity.

	T	5%	50%	95%	mean	Rel. MSE	5%	50%	95%	mean	Rel. MSE
			Case I						Case IV		
PDOLS	50	0.612	1.006	1.385	1.003	1.000	0.310	0.967	1.585	0.955	1.000
RDSUR	50	0.619	1.007	1.381	1.004	0.974	0.482	0.992	1.480	0.989	0.599
FRDSUR	50	0.584	1.005	1.427	1.004	1.182	0.427	0.986	1.513	0.982	0.733
PDOLS	100	0.807	0.999	1.190	0.999	1.000	0.680	0.974	1.256	0.974	1.000
RDSUR	100	0.810	0.999	1.191	1.000	0.976	0.761	0.991	1.223	0.991	0.640
FRDSUR	100	0.797	0.999	1.206	1.000	1.134	0.740	0.988	1.228	0.986	0.709
PDOLS	300	0.934	1.000	1.067	1.000	1.000	0.901	0.992	1.080	0.991	1.000
RDSUR	300	0.935	0.999	1.067	1.000	0.973	0.919	0.998	1.075	0.998	0.731
FRDSUR	300	0.933	1.000	1.069	1.000	1.044	0.919	0.996	1.073	0.996	0.712
			Case II						Case V		
PDOLS	50	0.643	1.009	1.368	1.007	1.000	0.487	0.981	1.428	0.975	1.000
RDSUR	50	0.654	1.008	1.361	1.007	0.944	0.633	1.009	1.375	1.010	0.623
FRDSUR	50	0.610	1.009	1.402	1.007	1.188	0.578	1.002	1.389	0.998	0.801
PDOLS	100	0.830	1.001	1.179	1.001	1.000	0.785	0.986	1.176	0.984	1.000
RDSUR	100	0.836	1.000	1.172	1.001	0.927	0.837	1.005	1.167	1.004	0.716
FRDSUR	100	0.821	1.001	1.191	1.002	1.130	0.826	0.999	1.160	0.997	0.756
PDOLS	300	0.944	1.001	1.060	1.001	1.000	0.938	0.996	1.050	0.995	1.000
RDSUR	300	0.947	1.001	1.057	1.001	0.902	0.952	1.003	1.055	1.002	0.812
FRDSUR	300	0.944	1.001	1.062	1.001	1.030	0.951	1.000	1.047	0.999	0.700
			Case III						Case VI		
PDOLS	50	0.619	1.022	1.403	1.019	1.000	0.441	0.990	1.503	0.984	1.000
RDSUR	50	0.768	1.026	1.288	1.028	0.460	0.761	1.022	1.290	1.023	0.265
FRDSUR	50	0.676	1.024	1.361	1.024	0.824	0.693	1.019	1.329	1.017	0.414
PDOLS	100	0.813	1.008	1.209	1.010	1.000	0.756	0.991	1.213	0.989	1.000
RDSUR	100	0.895	1.015	1.144	1.017	0.424	0.895	1.012	1.136	1.013	0.292
FRDSUR	100	0.867	1.015	1.170	1.016	0.626	0.884	1.009	1.137	1.010	0.328
PDOLS	300	0.938	1.002	1.074	1.003	1.000	0.928	0.997	1.064	0.996	1.000
RDSUR	300	0.966	1.005	1.051	1.007	0.415	0.970	1.005	1.045	1.006	0.327
FRDSUR	300	0.963	1.005	1.054	1.006	0.464	0.969	1.004	1.043	1.004	0.304

Note: PDOLS is panel DOLS, RDSUR is restricted DSUR and FRDSUR is feasible restricted DSUR. Rel. MSE is relative (to panel DOLS) mean square error.

Table 5: DSUR Estimation of Spot and Forward Exchange Rate Cointegrating Regression, 1975.1-1996.12

A. Leads and lags	3 leads and lags		2 leads and lags	
	$\hat{\beta}$	$t(\beta = 1)$	$\hat{\beta}$	$t(\beta = 1)$
Germany	0.992	-2.581	0.992	-2.191
Japan	1.000	0.247	1.000	0.199
UK	1.001	0.351	1.001	0.102
χ^2_2	7.459		5.135	
(p-value)	(0.024)		(0.077)	
χ^2_3	7.571		5.344	
(p-value)	(0.056)		(0.148)	
Restricted	0.997	-0.144	0.999	-0.271
B. Leads only	3 leads		2 leads	
	$\hat{\beta}$	$t(\beta = 1)$	$\hat{\beta}$	$t(\beta = 1)$
Germany	0.992	-1.860	0.992	-1.797
Japan	1.000	0.310	1.001	0.217
UK	1.001	0.271	1.000	0.031
χ^2_2	4.047		3.663	
(p-value)	(0.132)		(0.160)	
χ^2_3	4.064		3.721	
(p-value)	(0.254)		(0.293)	
Restricted	1.000	-0.047	1.000	-0.116

Notes: χ^2_2 is the test statistic for testing the homogeneity hypothesis $\beta_1 = \beta_2 = \beta_3$. χ^2_3 is the test statistic for testing the homogeneity hypothesis $\beta_1 = \beta_2 = \beta_3 = 1$.

Table 6: Saving-Investment Correlations

	Ratios		Log-Levels	
	$\hat{\beta}_i$	t($\beta_i = 1$)	$\hat{\beta}_i$	t($\beta_i = 1$)
A. European System				
Austria	1.071	0.486	1.021	1.050
Finland	1.408	4.636	0.859	-2.431
France	1.013	0.169	0.977	-0.885
Germany	0.762	-1.425	0.992	-0.116
Italy	1.211	3.014	0.965	-1.842
Spain	0.668	-2.024	0.981	-0.559
Switzerland	1.330	2.661	0.909	-3.250
UK	0.559	-2.882	0.986	-0.230
χ^2_7	29.10		4.487	
(p-value)	(0.000)		(0.722)	
χ^2_8	37.45		9.897	
(p-value)	(0.000)		(0.272)	
Restricted	—	—	0.974	-1.857
B. Non-European System				
Australia	0.600	-4.255	0.995	-0.139
Canada	0.818	-1.052	0.989	-0.183
Japan	0.974	-0.191	0.971	-1.208
US	0.878	-1.371	1.095	3.393
χ^2_3	3.771		2.421	
(p-value)	(0.287)		(0.490)	
χ^2_4	11.83		3.589	
(p-value)	(0.019)		(0.464)	
Restricted	0.777	-3.597	1.019	1.357

Note: Statistic for test of homogeneity is χ^2_7 in panel A and χ^2_3 in panel B. Statistic for test that slope coefficients are all equal to 1 is χ^2_8 in panel A and χ^2_4 in panel B.

Appendix

Proof of proposition 1. We note that three regularity conditions assumed by Saikkonen (1993) (i) the spectral density matrix of the vector of equilibrium errors is bounded away from zero, ii) the long-run covariance matrix exists, and iii) the 4-th order cumulants are absolutely sumable) are satisfied under assumption 1. Let $T_* = T - 2p$,

$$\mathbf{A} = \operatorname{diag}\left(\frac{1}{T_*^2}\left(\sum_{t=p+1}^{T-p} \mathbf{X}_t \mathbf{\Omega}_{uu}^{-1} \mathbf{X}_t'\right), \mathrm{E}\left(\mathbf{Z}_{pt} \mathbf{\Omega}_{uu}^{-1} \mathbf{Z}_{pt}'\right)\right), \qquad \mathbf{G}_T = \operatorname{diag}\left(T_* \mathbf{I}_2, \sqrt{T_*}\mathbf{I}_2\right)$$

and

$$\hat{\mathbf{A}} = \left[\mathbf{G}_T^{-1} \sum_{t=p+1}^{T-p} \left(\mathbf{W}_t \mathbf{\Omega}_{uu}^{-1} \mathbf{W}_t'\right) \mathbf{G}_T^{-1}\right] = \sum_{t=p+1}^{T-p} \begin{bmatrix} \frac{\Omega^{11}\underline{x}_{1t}\underline{x}_{1t}'}{T_*^2} & \frac{\Omega^{12}\underline{x}_{1t}\underline{x}_{2t}'}{T_*^2} & \frac{\Omega^{11}\underline{x}_{1t}\underline{z}_t'}{T_*^{3/2}} & \frac{\Omega^{12}\underline{x}_{1t}\underline{z}_t'}{T_*^{3/2}} \\ \frac{\Omega^{21}\underline{x}_{2t}\underline{x}_{1t}'}{T_*^2} & \frac{\Omega^{22}\underline{x}_{2t}\underline{x}_{2t}'}{T_*^2} & \frac{\Omega^{21}\underline{x}_{2t}\underline{z}_t'}{T_*^{3/2}} & \frac{\Omega^{22}\underline{x}_{2t}\underline{z}_t'}{T_*^{3/2}} \\ \frac{\Omega^{11}\underline{z}_t\underline{x}_{1t}'}{T_*^{3/2}} & \frac{\Omega^{12}\underline{z}_t\underline{x}_{2t}'}{T_*^{3/2}} & \frac{\Omega^{11}\underline{z}_t\underline{z}_t'}{T_*} & \frac{\Omega^{12}\underline{z}_t\underline{z}_t'}{T_*} \\ \frac{\Omega^{21}\underline{z}_t\underline{x}_{1t}'}{T_*^{3/2}} & \frac{\Omega^{22}\underline{z}_t\underline{x}_{2t}'}{T_*^{3/2}} & \frac{\Omega^{21}\underline{z}_t\underline{z}_t'}{T_*} & \frac{\Omega^{22}\underline{z}_t\underline{z}_t'}{T_*} \end{bmatrix}.$$

Then

$$\begin{aligned} \begin{bmatrix} T_*(\hat{\underline{\beta}}_{dsur} - \underline{\beta}) \\ \sqrt{T_*}(\hat{\underline{\delta}}_{p,dsur} - \underline{\delta}_p) \end{bmatrix} &= \hat{\mathbf{A}}^{-1}\mathbf{G}_T^{-1} \sum_{t=p+1}^{T-p} \mathbf{W}_t \mathbf{\Omega}_{uu}^{-1}(\underline{u}_t + \underline{v}_{pt}) \\ &= \mathbf{A}^{-1}\mathbf{G}_T^{-1} \sum_{t=p+1}^{T-p} \mathbf{W}_t \mathbf{\Omega}_{uu}^{-1}\underline{u}_t + \underbrace{\mathbf{A}^{-1}\mathbf{G}_T^{-1} \sum_{t=p+1}^{T-p} \mathbf{W}_t \mathbf{\Omega}_{uu}^{-1}\underline{v}_{pt}}_{(a)} \\ &\quad + \underbrace{\left(\hat{\mathbf{A}}^{-1} - \mathbf{A}^{-1}\right)\mathbf{G}_T^{-1} \sum_{t=p+1}^{T-p} \mathbf{W}_t \mathbf{\Omega}_{uu}^{-1}\left(\underline{v}_{pt} + \underline{u}_t\right)}_{(b)} \end{aligned}$$

From theorem 4.1 of Saikkonen (1993), we have $\mathbf{G}_T^{-1}\sum_{t=p+1}^{T-p} \mathbf{W}_t \mathbf{\Omega}_{uu}^{-1}\underline{v}_{pt} = o_p(1)$ and $\hat{\mathbf{A}}^{-1} - \mathbf{A}^{-1} = o_p(1)$ so that terms (a) and (b) above are both $o_p(1)$.
The block-diagonality of $\mathbf{A}^{-1}$ tells us that $T_*(\hat{\underline{\beta}}_{dsur} - \underline{\beta})$ and $\sqrt{T_*}(\hat{\underline{\delta}}_{p,dsur} - \underline{\delta}_p)$ are asymptotically independent. It follows that

$$\begin{aligned} T_*(\hat{\underline{\beta}}_{dsur} - \underline{\beta}) &= \left(\frac{1}{T_*^2}\sum \mathbf{X}_t \mathbf{\Omega}_{uu}^{-1}\mathbf{X}_t\right)^{-1}\left(\frac{1}{T_*}\sum \mathbf{X}_t \mathbf{\Omega}_{uu}^{-1}\underline{u}_t\right) + o_p(1) \\ &\xrightarrow{D} \left(\int \mathbf{B}_e \mathbf{\Omega}_{uu}^{-1}\mathbf{B}_e'\right)^{-1}\left(\int \mathbf{B}_e \mathbf{\Omega}_{uu}^{-1} d\underline{B}_u'\right) \end{aligned} \tag{A.1}$$

Conditional on $\mathbf{B}_e$, $\int \mathbf{B}_e \mathbf{\Omega}_{uu}^{-1} d\underline{B}_u \xrightarrow{D} \mathrm{N}(\underline{0}, [\int \mathbf{B}_e \mathbf{\Omega}_{uu}^{-1}\mathbf{B}_e'])$ [Park and Phillips (1998)].

Let $\mathbf{R}$ be a $q \times 2k$ restriction matrix. Note that $\mathbf{B}_e$ and $\underline{B}_u$ are independent Brownian motions. Then conditional on $\mathbf{B}_e$,

$$(\mathbf{R}(\underline{\hat{\beta}}_{dsur} - \underline{\beta})'[\mathbf{R}(\int \mathbf{B}_e \mathbf{\Omega}_{uu}^{-1} \mathbf{B}_e')\mathbf{R}']^{-1}(\mathbf{R}(\underline{\hat{\beta}}_{dsur} - \underline{\beta}) \xrightarrow{D} \chi_q^2. \tag{A.2}$$

Since the chi-square distribution does not depend on $\int \mathbf{B}_e \mathbf{\Omega}_{uu}^{-1} \mathbf{B}_e'$, and $\frac{1}{T^2}\sum_t^T \mathbf{X}_t \mathbf{\Omega}_{uu}^{-1} \mathbf{X}_t' \xrightarrow{D} \int \mathbf{B}_e \mathbf{\Omega}_{uu}^{-1} \mathbf{B}_e'$, a test of the null hypothesis $H_o : \mathbf{R}\underline{\hat{\beta}}_{dsur} = \underline{r}$, can be conducted with the Wald statistic

$$(\mathbf{R}\underline{\hat{\beta}}_{dsur} - \underline{r})' \left[\mathbf{R}\left(\sum_{t=1}^{T} \mathbf{X}_t \mathbf{\Omega}_{uu}^{-1} \mathbf{X}_t'\right)\mathbf{R}\right]^{-1} (\mathbf{R}\underline{\hat{\beta}}_{dsur} - \underline{r}) \tag{A.3}$$

which has a limiting χ_q^2 distribution. $\|$

To prove proposition 2, we make use of the following lemma.

Lemma 1 The two-step OLS-SUR estimator is numerically equivalent to the two-step GLS-SUR estimator.

Proof. Let $\underline{Y} = \left(\underline{y}_1', \underline{y}_2'\right)$, $\underline{y}_i' = (y_{i,p+1}, \ldots, y_{i,T-p})$, $\mathbf{X} = \text{diag}\,(\mathbf{x}_1, \mathbf{x}_2)$, $\mathbf{x}_i = \left(\underline{x}_{i,p+1}, \ldots, \underline{x}_{i,T-p}\right)'$, $\mathbf{Z} = \text{diag}\,(\mathbf{z}_p, \mathbf{z}_p) = (\mathbf{I}_2 \otimes \mathbf{z}_p)$, $\mathbf{z}_p = \left(\underline{z}_{p,p+1}, \ldots, \underline{z}_{p,T-p}\right)'$, $\underline{\beta} = \left(\underline{\beta}_1', \underline{\beta}_2'\right)'$, $\underline{\delta}_p = \left(\underline{\delta}_{p1}', \underline{\delta}_{p2}'\right)'$, $\underline{U} = (\underline{u}_1, \underline{u}_2)$, $\underline{u}_i = (u_{i,p+1}, \ldots, u_{i,T-p})$. Write (7) in matrix form,

$$\underline{Y} = \mathbf{X}\underline{\beta} + \mathbf{Z}\underline{\delta}_p + \underline{U} \tag{A.4}$$

Let $\mathbf{M} = \mathbf{I} - \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'$, $\mathbf{\Omega}_{uu}^{-1} = \mathbf{P}'\mathbf{P}, \mathbf{H} = \mathbf{P} \otimes \mathbf{I}$, and $\mathbf{V} = \mathbf{H}'\mathbf{H} = (\mathbf{\Omega}^{-1} \otimes \mathbf{I})$ (note: $\mathbf{P} = \mathbf{L}^{-1}$ in the text). Then $\mathbf{M}\underline{Y}$ is the vector of OLS residuals from regressing y_{it} on $\underline{z}_{pt}$ and $\mathbf{MX}$ is the corresponding matrix of OLS residuals from regressing $\underline{x}_{it}$ on $\underline{z}_{pt}$. The two-step OLS-SUR estimator is obtained from applying OLS to $\mathbf{HM}\underline{Y} = \mathbf{HMX}\underline{\beta} + \mathbf{HM}\underline{U}$. which gives

$$\underline{\hat{\beta}}_A = \left(\mathbf{X}'\mathbf{M}'\mathbf{V}\mathbf{M}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{M}'\mathbf{V}\mathbf{M}\underline{Y}\right).$$

To obtain the two-step GLS-SUR estimator, premultiply (A.4) by H to obtain $\underline{Y}_* = \mathbf{X}_*\underline{\beta} + \mathbf{Z}_*\underline{\delta} + \underline{U}_*$, where $\underline{Y}_* = \mathbf{H}\underline{Y}$, $\mathbf{Z}_* = \mathbf{HZ}$, $\underline{U}_* = \mathbf{H}\underline{U}$. Let $\mathbf{M}_* = \mathbf{I} - \mathbf{Z}_*\left(\mathbf{Z}_*'\mathbf{Z}_*\right)^{-1}\mathbf{Z}_*'$. Then $\mathbf{M}_*\underline{Y}_*$ is the vector of GLS residuals from regressing y_{it} on $\underline{z}_{pt}$, and $\mathbf{M}_*\mathbf{X}_*$ is the corresponding matrix of GLS residuals from regressing $\underline{x}_{it}$ on $\underline{z}_{pt}$. The two-step GLS-SUR estimator is obtained by applying OLS to $\mathbf{M}_*\underline{Y}_* = \mathbf{M}_*\mathbf{X}_*\underline{\beta} + \mathbf{M}_*\underline{U}_*$, which gives

$$\underline{\hat{\beta}}_B = \left(\mathbf{X}_*'\mathbf{M}_*\mathbf{X}_*\right)^{-1}\left(\mathbf{X}_*'\mathbf{M}_*\underline{Y}_*\right).$$

Noting that $\mathbf{VZ} = (\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{z}_p)$ and $\mathbf{Z'Z} = \left(\mathbf{I} \otimes \mathbf{z}_p'\mathbf{z}_p\right)$, it is straightforward to see that $\underline{\hat{\beta}}_A = \underline{\hat{\beta}}_B$. ‖

Proof of proposition 2. In addition to the matrix notation developed for lemma 1, let $\underline{V}_p = \left(\underline{v}_{p1}, \underline{v}_{p2}\right)'$, $\underline{v}_{pi} = (v_{pi,p+1}, \ldots, v_{pi,T-p})'$, and $\underline{V}_{*p} = H\underline{V}_p$ We have for the two-step GLS-SUR estimator of $\underline{\beta}$,

$$T_* \left(\underline{\hat{\beta}} - \underline{\beta}\right) = \Bigg(\underbrace{\frac{1}{T_*^2}\mathbf{X}_*'\mathbf{M}_*\mathbf{X}_*}_{E_1}\Bigg)^{-1} \Bigg(\underbrace{\frac{1}{T_*}\mathbf{X}_*'\mathbf{M}_*\left(\underline{U}_* + \underline{V}_{*p}\right)}_{E_2}\Bigg)$$

For E_2,

$$\begin{aligned}
\frac{1}{T_*}\mathbf{X}_*'\mathbf{M}_*\left(\underline{U}_* + \underline{V}_{*p}\right) &= \frac{1}{T_*}\mathbf{X}'\mathbf{H}'\left(\mathbf{I} - \mathbf{HZ}\left(\mathbf{Z}'\mathbf{VZ}\right)^{-1}\mathbf{Z}'\mathbf{H}'\right)\mathbf{H}\left(\underline{U} + \underline{V}_p\right) \\
&= \underbrace{\frac{1}{T_*}\mathbf{X}'\left(\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{I}\right)\left(\underline{U} + \underline{V}_p\right)}_{(a)} \\
&\qquad - \underbrace{\frac{1}{T_*}\mathbf{X}'\left(\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{z}_p\left(\mathbf{z}_p'\mathbf{z}_p\right)^{-1}\mathbf{z}_p\right)\left(\underline{U} + \underline{V}_p\right)}_{(b)}
\end{aligned}$$

For term (a),

$$\begin{aligned}
\frac{1}{T_*}\mathbf{X}'\left(\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{I}\right)\left(\underline{U} + \underline{V}_p\right) &= \frac{1}{T_*}\begin{bmatrix} \sum_{t=p+1}^{T-p}\underline{x}_{1t}\left(\Omega_{uu}^{11}(u_{1t} + v_{p1t}) + \Omega_{uu}^{12}(u_{2t} + v_{p2t})\right) \\ \sum_{t=p+1}^{T-p}\underline{x}_{2t}\left(\Omega_{uu}^{21}(u_{1t} + v_{p1t}) + \Omega_{uu}^{22}(u_{2t} + v_{p2t})\right) \end{bmatrix} \\
&= \frac{1}{T_*}\begin{bmatrix} \sum_{t=p+1}^{T-p}\underline{x}_{1t}\left(\Omega_{uu}^{11}u_{1t} + \Omega_{uu}^{12}u_{2t}\right) \\ \sum_{t=p+1}^{T-p}\underline{x}_{2t}\left(\Omega_{uu}^{21}u_{1t} + \Omega_{uu}^{22}u_{2t}\right) \end{bmatrix} + o_p(1) \\
&\xrightarrow{D} \int \mathbf{B}_e\mathbf{\Omega}_{uu}^{-1}\mathbf{B}_e'
\end{aligned}$$

For term (b),

$$\begin{aligned}
&\frac{1}{T_*}\mathbf{X}'\left(\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{z}_p\left(\mathbf{z}_p'\mathbf{z}_p\right)^{-1}\mathbf{z}_p\right)\left(\underline{U} + \underline{V}_p\right) \\
&\quad = \frac{1}{T_*}\begin{bmatrix} \Omega_{uu}^{11}\mathbf{x}_1'\mathbf{z}_p\left(\mathbf{z}_p'\mathbf{z}_p\right)^{-1}\mathbf{z}_p(\underline{u}_1 + \underline{v}_{p1}) + \Omega_{uu}^{12}\mathbf{x}_1'\mathbf{z}_p\left(\mathbf{z}_p'\mathbf{z}_p\right)^{-1}\mathbf{z}_p(\underline{u}_2 + \underline{v}_{p2}) \\ \Omega_{uu}^{21}\mathbf{x}_2'\mathbf{z}_p\left(\mathbf{z}_p'\mathbf{z}_p\right)^{-1}\mathbf{z}_p(\underline{u}_1 + \underline{v}_{p1}) + \Omega_{uu}^{22}\mathbf{x}_2'\mathbf{z}_p\left(\mathbf{z}_p'\mathbf{z}_p\right)^{-1}\mathbf{z}_p(\underline{u}_2 + \underline{v}_{p2}) \end{bmatrix} \\
&\quad = \begin{bmatrix} \sum_{j=1}^{2}\Omega_{uu}^{1j}\left(\frac{1}{T_*}\sum_{t=p+1}^{T-p}\underline{x}_{1t}\underline{z}_{pt}'\right)\left(\frac{1}{T_*}\sum_{t=p+1}^{T-p}\underline{z}_{pt}\underline{z}_{pt}'\right)\left(\frac{1}{T_*}\sum_{t=p+1}^{T-p}\underline{z}_{pt}(u_{jt} + v_{pjt})\right) \\ \sum_{j=1}^{2}\Omega_{uu}^{2j}\left(\frac{1}{T_*}\sum_{t=p+1}^{T-p}\underline{x}_{2t}\underline{z}_{pt}'\right)\left(\frac{1}{T_*}\sum_{t=p+1}^{T-p}\underline{z}_{pt}\underline{z}_{pt}'\right)\left(\frac{1}{T_*}\sum_{t=p+1}^{T-p}\underline{z}_{pt}(u_{jt} + v_{pjt})\right) \end{bmatrix}
\end{aligned}$$

$$= \begin{bmatrix} o_p(1) \\ o_p(1) \end{bmatrix}$$

It follows that $E_2 = \frac{1}{T_*}\mathbf{X}_*'\mathbf{M}_*\left(\underline{U}_* + \underline{V}_{*p}\right) \xrightarrow{D} \int \mathbf{B}_e \mathbf{\Omega}_{uu}^{-1} d\underline{B}_u$.
Next, we have for E_1,

$$\frac{1}{T_*^2}\mathbf{X}_*'\mathbf{M}_*\mathbf{X}_* = \underbrace{\frac{1}{T_*^2}\mathbf{X}'\mathbf{V}\mathbf{X}}_{(c)} - \underbrace{\frac{1}{T_*^2}\mathbf{X}'\left(\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{z}_p\left(\mathbf{z}_p'\mathbf{z}_p\right)^{-1}\mathbf{z}_p'\right)\mathbf{X}}_{(d)}$$

Expanding term (d) gives

$$\begin{aligned} \frac{1}{T_*^2}\mathbf{X}'\left(\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{z}_p\left(\mathbf{z}_p'\mathbf{z}_p\right)^{-1}\mathbf{z}_p'\right)\mathbf{X} &= \frac{1}{T_*^2}\begin{bmatrix} \Omega_{uu}^{11}\mathbf{x}_1'\mathbf{z}_p(\mathbf{z}_p'\mathbf{z}_p)^{-1}\mathbf{z}_p\mathbf{x}_1 & \Omega_{uu}^{12}\mathbf{x}_1'\mathbf{z}_p(\mathbf{z}_p'\mathbf{z}_p)^{-1}\mathbf{z}_p\mathbf{x}_2 \\ \Omega_{uu}^{21}\mathbf{x}_2'\mathbf{z}_p(\mathbf{z}_p'\mathbf{z}_p)^{-1}\mathbf{z}_p\mathbf{x}_1 & \Omega_{uu}^{22}\mathbf{x}_2'\mathbf{z}_p(\mathbf{z}_p'\mathbf{z}_p)^{-1}\mathbf{z}_p\mathbf{x}_2 \end{bmatrix} \\ &= o_p(1) \end{aligned}$$

since the ij-th element of the matrix is $\Omega^{ij}\left(\frac{1}{T_*^{3/2}}\sum_{t=p+1}^{T-p}\underline{x}_{it}\underline{z}_{pt}'\right)\left(\frac{1}{T_*}\sum_{t=p+1}^{T-p}\underline{z}_{pt}\underline{z}_{pt}'\right)\left(\frac{1}{T_*^{3/2}}\sum_{t=p+1}^{T-p}\underline{z}_{pt}\underline{x}_{jt}'\right)$ $= o_p(1)$.

Expanding term (c) gives

$$\frac{1}{T_*^2}\mathbf{X}'\mathbf{V}\mathbf{X} = \frac{1}{T_*^2}\begin{bmatrix} \Omega_{uu}^{11}\sum_{t=p+1}^{T-p}\underline{x}_{1t}\underline{x}_{1t}' & \Omega_{uu}^{12}\sum_{t=p+1}^{T-p}\underline{x}_{1t}\underline{x}_{2t}' \\ \Omega_{uu}^{21}\sum_{t=p+1}^{T-p}\underline{x}_{2t}\underline{x}_{1t}' & \Omega_{uu}^{22}\sum_{t=p+1}^{T-p}\underline{x}_{2t}\underline{x}_{2t}' \end{bmatrix} \xrightarrow{D} \int \mathbf{B}_e\mathbf{\Omega}_{uu}^{-1}\mathbf{B}_e$$

Thus, it is established that $E_1 = \frac{1}{T_*^2}\mathbf{X}_*'\mathbf{M}_*\mathbf{X}_* \xrightarrow{D} \int \mathbf{B}_e\mathbf{\Omega}_{uu}^{-1}\mathbf{B}_e$. By lemma 1, the equivalence of the OLS-SUR two-step estimator and DSUR obtains. $\|$

Proof of proposition 3. $T_*\left(\hat{\underline{\beta}}_{sysdols} - \underline{\beta}\right) = \left(\frac{1}{T_*^2}\sum_{t=p+1}^{T-p}\tilde{\mathbf{X}}_t\tilde{\mathbf{X}}_t'\right)^{-1}\left(\frac{1}{T_*}\sum_{t=p+1}^{T-p}\tilde{\mathbf{X}}_t\underline{u}_t\right)$. From proposition 1 we have $\frac{1}{T_*^2}\sum_{t=p+1}^{T-p}\tilde{\mathbf{X}}_t\tilde{\mathbf{X}}_t' \xrightarrow{D} \int \mathbf{B}_e\mathbf{B}_e' = \text{diag}\left(\int \underline{B}_{e_1}\underline{B}_{e_1}', \int \underline{B}_{e_2}\underline{B}_{e_2}'\right)$, and $\frac{1}{T_*}\sum_{t=p+1}^{T-p}\tilde{\mathbf{X}}_t\underline{u}_t \xrightarrow{D} \int \mathbf{B}_e d\underline{B}_u' = \left(\int \mathbf{B}_{e_1}' d\underline{B}_{u_1}, \int \mathbf{B}_{e_2}' d\underline{B}_{u_2}\right)'$. Conditional on $\mathbf{B}_e$, $T\left(\hat{\underline{\beta}}_{sysdols} - \underline{\beta}\right) \sim N(0, \mathbf{V}_{sysdols})$ where $\mathbf{V}_{sysdols} = \left(\int \mathbf{B}_e\mathbf{B}_e'\right)^{-1}\left(\int \mathbf{B}_e\mathbf{\Omega}_{uu}\mathbf{B}_e'\right)\left(\int \mathbf{B}_e\mathbf{B}_e'\right)^{-1}$. The asymptotic chi-square distribution of the Wald statistic follows immediately from the mixed-normality of the estimator.

To prove proposition 4, we make use of the following two lemmas.

Lemma 2

$$\text{avar}(\hat{\underline{\beta}}_{dsur}) = \mathsf{E}\left(\int \mathbf{B}_e\mathbf{\Omega}_{uu}^{-1}\mathbf{B}_e'\right)^{-1}$$

$$\text{avar}(\underline{\hat{\beta}}_{sysdols}) = \text{E}\left(\int \mathbf{B}_e\mathbf{B}_e'\right)^{-1}\left(\int \mathbf{B}_e\mathbf{\Omega}_{uu}\mathbf{B}_e'\right)\left(\int \mathbf{B}_e\mathbf{B}_e'\right)^{-1}.$$

Proof. Conditional on $\mathbf{B}_e$, $\text{avar}(\underline{\hat{\beta}}_{dsur}) = \mathbf{V}_1^{-1}$, where $\mathbf{V}_1 = \int \mathbf{B}_e\mathbf{\Omega}_{uu}^{-1}\mathbf{B}_e'$. It follows that

$$\begin{aligned}
\text{Var}\left[\mathbf{V}_1^{-1}\int \mathbf{B}_e\mathbf{\Omega}_{uu}^{-1}d\underline{B}_u|\mathbf{B}_e\right] &= \mathbf{V}_1^{-1}\left(\int \mathbf{B}_e\mathbf{\Omega}_{uu}^{-1}\mathbf{\Omega}_{uu}\mathbf{\Omega}_{uu}^{-1}\mathbf{B}_e'\right)\mathbf{V}_1^{-1} = \mathbf{V}_1^{-1} \\
\text{E}\left[\mathbf{V}^{-1}\int \mathbf{B}_e\mathbf{\Omega}_{uu}^{-1}d\underline{B}_u|\mathbf{B}_e\right] &= 0
\end{aligned}$$

Using the decomposition of the variance for any two random variables Y and X,

$$\text{Var}(Y) = \text{E}\left[\text{Var}(Y|X)\right] + \text{Var}\left[\text{E}(Y|X)\right], \tag{A.5}$$

it follows that unconditionally, $\text{avar}(\underline{\hat{\beta}}_{dsur}) = \text{E}\left(\text{Var}\left(\mathbf{V}_1^{-1}\int \mathbf{B}_e\mathbf{\Omega}_{uu}^{-1}d\underline{B}_u\right)\right) = \text{E}\left(\int \mathbf{B}_e\mathbf{\Omega}_{uu}^{-1}\mathbf{B}_e'\right)^{-1}$. Similarly, we have $\text{avar}(\underline{\hat{\beta}}_{sysdols}) = \text{E}\left(\left(\int \mathbf{B}_e\mathbf{B}_e'\right)^{-1}\left(\int \mathbf{B}_e\mathbf{\Omega}_{uu}\mathbf{B}_e'\right)\left(\int \mathbf{B}_e\mathbf{B}_e'\right)^{-1}\right)$. ‖

Lemma 3 Consider the random matrices $\mathbf{A}_T$ and $\mathbf{B}_T$. If $\mathbf{A}_T \geq \mathbf{B}_T$, $\mathbf{A}_T \xrightarrow{D} \mathbf{A}$ and $\mathbf{B}_T \xrightarrow{D} \mathbf{B}$, then $\mathbf{A} \geq \mathbf{B}$, almost surely.

Proof. Given $\underline{\lambda}'(\mathbf{A}_T - \mathbf{B}_T)\underline{\lambda} \geq 0$. Assume the converse: $\text{P}\left(\underline{\lambda}'\left(\mathbf{A}-\mathbf{B}\right)\underline{\lambda} < 0\right) > 0$. Then there exists an $\epsilon > 0$ such that $\text{P}\left(\underline{\lambda}'\left(\mathbf{A}-\mathbf{B}\right)\underline{\lambda} < -\epsilon\right) > 0$. There are a countable number of continuity points within the interval $[-\epsilon, 0]$. Let $-\delta$ be one such continuity point where, $-\epsilon < -\delta < 0$. Then $\lim_T \text{P}\left(\underline{\lambda}'\left(\mathbf{A}_T-\mathbf{B}_T\right)\underline{\lambda} < -\delta\right) = \text{P}\left(\underline{\lambda}'\left(\mathbf{A}-\mathbf{B}\right)\underline{\lambda} < -\delta\right) > 0$, which is a contradiction. ‖

Proof of proposition 4. Let

$$\mathbf{x}_t = \text{diag}\left(\underline{x}_{1t}, \underline{x}_{2t}\right) : (2k \times 2), \qquad \mathbf{X}_T = \text{diag}\left(\mathbf{x}_{p+1}, \ldots, \mathbf{x}_{T-p}\right) : (2T_*k \times 2T_*)$$

$$\mathbf{V}_{1T_*} = \frac{\mathbf{X}_{T_*}'\mathbf{\Omega}^{-1}\mathbf{X}_{T_*}}{T_*^2} \qquad \mathbf{V}_{2T_*} = \left(\frac{\mathbf{X}_{T_*}'\mathbf{X}_{T_*}}{T_*^2}\right)\left(\frac{\mathbf{X}_{T_*}'\mathbf{\Omega}\mathbf{X}_{T_*}}{T_*^2}\right)^{-1}\left(\frac{\mathbf{X}_{T_*}'\mathbf{X}_{T_*}}{T_*^2}\right)$$

Then

$$\begin{aligned}
\mathbf{V}_{1T_*} - \mathbf{V}_{2T_*} &= \frac{\mathbf{X}_{T_*}'\mathbf{\Omega}^{-1}\mathbf{X}_{T_*}}{T_*^2} - \left(\frac{\mathbf{X}_{T_*}'\mathbf{X}_{T_*}}{T_*^2}\right)\left(\frac{\mathbf{X}_{T_*}'\mathbf{\Omega}\mathbf{X}_{T_*}}{T_*^2}\right)^{-1}\left(\frac{\mathbf{X}_{T_*}'\mathbf{X}_{T_*}}{T_*^2}\right) \\
&= \left(\frac{\mathbf{X}_{T_*}'\mathbf{\Omega}^{-1/2}}{T_*}\right)\left[\mathbf{I} - \left(\frac{\mathbf{\Omega}^{1/2}\mathbf{X}_{T_*}}{T_*}\right)\left(\frac{\mathbf{X}_{T_*}'\mathbf{\Omega}^{1/2}\mathbf{\Omega}^{1/2}\mathbf{X}_{T_*}}{T_*^2}\right)\left(\frac{\mathbf{X}_{T_*}'\mathbf{\Omega}^{1/2}}{T_*}\right)\right]\left(\frac{\mathbf{\Omega}^{-1/2}\mathbf{X}_{T_*}}{T_*}\right) \\
&= \mathbf{D}_{T_*}'\left[\mathbf{I} - \mathbf{M}_{T_*}\left(\mathbf{M}_{T_*}'\mathbf{M}_{T_*}\right)^{-1}\mathbf{M}_{T_*}'\right]\mathbf{D}_{T_*}
\end{aligned}$$

where $\mathbf{D}_{T_*} = (1/T_*)\mathbf{\Omega}^{-1/2}\mathbf{X}_{T_*} : (2T_* \times 2T_*)$ and $\mathbf{M}_{T_*} = (1/T_*)\mathbf{\Omega}^{1/2}\mathbf{X}_{T_*}$. This is a system of $2T_*$ nonnegative quadratic forms in a symmetric idempotent matrix. For given $\mathbf{X}_{T_*}$ and T_*, we have $\mathbf{V}_{1T_*} \geq \mathbf{V}_{2T_*}$ which implies that $\mathbf{V}_{1T_*}^{-1} \leq \mathbf{V}_{2T_*}^{-1}$.

By lemma 3, we have $\mathbf{V}_1^{-1} \leq \mathbf{V}_2^{-1}$, and lemma 2 gives $\text{avar}(\underline{\hat{\beta}}_{dsur}) \leq \text{avar}(\underline{\hat{\beta}}_{sysdols})$. $\|$