Review of Economic Studies (2005) **72**, 797–820 0034-6527/05/00320797$02.00


# Dynamic Seemingly Unrelated Cointegrating Regressions

NELSON C. MARK
*University of Notre Dame and NBER*

MASAO OGAKI
*Ohio State University*

and

DONGGYU SUL
*University of Auckland*

*First version received April* 2002; *final version accepted September* 2004 (*Eds.*)

We propose the parametric Dynamic Seemingly Unrelated Regression (DSUR) estimator for simultaneous estimation of multiple cointegrating regressions. DSUR is efficient when the equilibrium errors are correlated across equations and is applicable for panel cointegration estimation in environments where the cross section is small relative to the available time series. We study the asymptotic and small sample properties of the DSUR estimator for both heterogeneous and homogeneous cointegrating vectors. We then apply the method to analyse two long-standing problems in international economics. Our first application revisits the estimation of long-run correlations between national investment and national saving. Our second application revisits the question of whether the forward exchange rate is an unbiased predictor of the future spot rate.

## INTRODUCTION

Multiple-equation cointegrating regressions are frequently encountered in applied research. Many applications arise in the context of panel cointegration regression. For example, one might combine multiple macroeconomic time series from a cross section of countries to estimate long-run money demand elasticities, the relation between investment and saving shares, relations among asset prices or relations among commodity prices. In this paper, we propose a parametric method for estimating multiple cointegrating regressions called the Dynamic Seemingly Unrelated Regression (DSUR) estimator. The DSUR method is feasible for balanced panels where the number of cointegrating regression equations $N$ is substantially smaller than the number of time-series observations $T$ and is applicable both in environments where the cointegrating vectors are homogeneous across equations and where they are not. DSUR achieves significant efficiency gains over non-system methods such as dynamic ordinary least squares (DOLS) when heterogeneous sets of regressors enter into the regressions and when equilibrium errors are correlated across cointegrating regressions.

We illustrate the usefulness of DSUR by revisiting two long-standing problems in international economics. Our first application revisits the estimation of long-run national saving and investment correlations originally put forward by Feldstein and Horioka (1980). They study a cross-sectional regression of the time-series averages of national investment shares on national saving shares and reason that the estimated slope coefficient is inversely related to the degree of capital mobility. Finding the slope coefficient to be insignificantly different from 1, they conclude that the degree of international capital mobility is low. Coakley, Kulasi and Smith (1996) extend this work by showing that under a time-series interpretation, the solvency constraint restricts the current account balance (savings minus investment) to being stationary irrespective of the degree of capital mobility. They suggest that Feldstein and Horioka's cross-section regression may be capturing this long-run relationship when long-run time-series averages are used in the regression. By employing DSUR, we obtain efficient estimates of the cointegrating coefficients in a system of cointegrating regressions of national investment variables on saving variables for a small panel of OECD countries. This allows us to conduct a direct test of the long-run relationship implied by the solvency conditions by testing the hypothesis that the slope coefficient is 1. The weight of the evidence supports the hypothesis that the solvency condition is not violated.

Our second application revisits the Evans and Lewis (1995) analysis of cointegrating regressions of the future spot exchange rate on the current forward exchange rate. Finding the slope coefficient in this cointegrating regression to be significantly different from 1, they report a new anomaly in international finance—that the expected excess return from forward foreign exchange speculation is unit-root non-stationary. While Evans and Lewis employ an SUR cointegration vector estimator, their control for endogeneity is incomplete. When we update Evans and Lewis's sample and complete the endogeneity control with DSUR, we find the evidence for a non-stationary expected excess return to be less compelling.

DSUR provides a parametric alternative to non-parametric estimators of seemingly unrelated cointegrating regressions proposed by Park and Ogaki (1991), who generalized the Park (1992) Canonical Cointegrating Regression estimators and by Moon (1999) who generalized the Phillips and Hansen (1990) fully modified estimators.[1] DSUR estimators are asymptotically equivalent to these non-parametric estimators. In finite samples, DSUR has the usual advantages and disadvantages compared to the non-parametric estimators: DSUR is more efficient than the non-parametric estimators if the parametric assumptions are correct, while the non-parametric methods are more robust.

We discuss the asymptotic properties of DSUR for $T \to \infty$ with $N$ fixed. For the estimation of heterogeneous cointegration vectors, we discuss the advantages of DSUR in relation to the following parametric estimators: DOLS, proposed by Phillips and Loretan (1991) and Stock and Watson (1993), a generalized DOLS estimator developed by Saikkonen (1991) which Park and Ogaki (1991) call "system DOLS", and a system estimator introduced by Saikkonen (1993). System DOLS is distinguished from ordinary DOLS 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 from all other equations in the system. Saikkonen (1991) developed system DOLS in a multivariate regression framework and showed that the system DOLS estimator is asymptotically more efficient than DOLS. The Saikkonen (1993) system estimator (SSE) is primarily intended for estimating "structural" coefficients in a system of cointegrating relations where linear identifying restrictions are available. This contrasts with DSUR which is primarily a strategy for estimating "reduced form" coefficients. However, the two estimators are comparable when the structure is identical to the reduced form. In this case, they are asymptotically equivalent but DSUR is efficient relative to SSE in finite samples.

Cross-equation restrictions (*e.g.* homogeneity restrictions) can be conveniently tested using Wald statistics which are asymptotically distributed as chi-square variates. If the null hypothesis of cointegration vector homogeneity is not rejected, estimation can be performed using a pooled

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

estimator of the cointegration vector that exploits the long-run dependence across individuals. We call this pooled estimator restricted DSUR and show that it is asymptotically efficient relative to panel DOLS.[2]

We conduct a series of Monte Carlo experiments to assess the small sample performance of DSUR in relation to alternative parametric estimators and the accuracy of the asymptotic theory. Because the equilibrium error may be correlated with an infinite number of leads and lags of changes in the regressors, a practical question of interest concerns the performance of parametric cointegration regression estimators that control for endogeneity by including a finite number of leads and lags. We find that the asymptotic distribution theory developed for DSUR works reasonably well and that there are important and sizable efficiency gains to be enjoyed by using DSUR over the DOLS methods and Saikkonen's SSE estimator.

The plan of the paper is as follows. Section 1 presents the estimator and discusses computational issues. We include in our discussion estimation of the required long-run covariance matrices and selection of lead and lag length. Section 2 contains a discussion of the asymptotic and small sample properties of the estimator and comparisons to alternative parametric estimators. Readers who are primarily interested in computational aspects of DSUR can skip this section without loss of continuity. The applications are presented in Section 3 and Section 4 concludes the paper. Proofs of propositions are contained in the Appendix.

## 1. THE DSUR ESTIMATOR

We consider a fixed number of $N$ cointegrating regressions each with $T$ observations. For example, the data may be balanced panels of individuals indexed by $i = 1, \ldots, N$ tracked over time periods $t = 1, \ldots, T$. Vectors are underlined and matrices appear in bold face but scalars have no special notation. The data are generated according to:

**Assumption 1 (Triangular Representation).** *Each equation $i = 1, \ldots, N$ 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. Letting $\underline{u}_t^{\dagger} = (u_{1t}^{\dagger}, \ldots, u_{NT}^{\dagger})'$, $\underline{e}_t' = (\underline{e}_{1t}', \ldots, \underline{e}_{NT}')'$, we have that $\underline{w}_t^{\dagger} = (\underline{u}_t^{\dagger'}, \underline{e}_t')'$ is an $N(k+1)$-dimensional vector with the orthonormal Wold moving average representation, $\underline{w}_t^{\dagger} = \mathbf{\Psi}^{\dagger}(L)\underline{\epsilon}_t$, where $\sum_{r=0}^{\infty} r|\psi_{ir}^{\dagger mn}| < \infty$, $\psi_{ir}^{\dagger mn}$ is the $m, n$-th element of the matrix $\mathbf{\Psi}_{ir}^{\dagger}$ and $\underline{\epsilon}_t$ is a martingale difference sequence with $E(\underline{\epsilon}_t) = \underline{0}$, $E(\underline{\epsilon}_t\underline{\epsilon}_t') = \mathbf{I}_k$, and finite fourth moments.*

The endogeneity problem shows up as correlation between the $i$-th equilibrium error $u_{it}^{\dagger}$ and potentially an infinite number of leads and lags of the first-differenced regressors from all of the equations in the system $\Delta\underline{x}_{jt} = \underline{e}_{jt}$, $(i, j = 1, \ldots, N)$. To control for endogeneity, we include leads and lags of these variables in the regressions. However, any feasible parametric estimation

2. Kao and Chiang (2000) and Mark and Sul (2003) studied the properties of panel DOLS under the assumption of independence across cross-sectional units. Phillips and Moon (1999) and Pedroni (2000) 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. While extant analyses of panel DOLS have been conducted under the assumption of independence across cross-sectional units, we show that the asymptotic distribution of panel DOLS is straightforward to obtain under cross-sectional dependence.

strategy can include only a finite number $p$ of leads and lags which induces a separate truncation error. To keep track of the truncation error, let

$$\begin{aligned}
\underline{z}'_{pit} &= (\Delta \underline{x}'_{it-p}, \ldots, \Delta \underline{x}'_{it+p}), \\
\underline{z}'_{pt} &= (\underline{z}'_{p1t}, \ldots, \underline{z}'_{pNt}), \\
\underline{\delta}_{p1} &= (\underline{\delta}'_{11,-p}, \ldots, \underline{\delta}'_{11,p}, \ldots, \underline{\delta}'_{1N,-p}, \ldots, \underline{\delta}'_{1N,p}), \\
&\vdots \\
\underline{\delta}_{pN} &= (\underline{\delta}'_{N1,-p}, \ldots, \underline{\delta}'_{N1,p}, \ldots, \underline{\delta}'_{NN,-p}, \ldots, \underline{\delta}'_{NN,p}),
\end{aligned}$$

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

$$u^{\dagger}_{it} = \underline{z}'_{pt}\underline{\delta}_{pi} + v_{pit} + u_{it}, \tag{3}$$

where

$$v_{pit} = \sum\nolimits_{j>|p|} \underline{\delta}'_{i1,j} \Delta \underline{x}_{it-j} + \cdots + \sum\nolimits_{j>|p|} \underline{\delta}'_{iN,j} \Delta \underline{x}_{N,t-j}, \tag{4}$$

are the truncation errors induced for given $p$ arising from the dependence of the equilibrium errors on $(\Delta \underline{x}'_{1t}, \ldots, \Delta \underline{x}'_{Nt})$ at distant leads and lags. Because the equilibrium errors $u^{\dagger}_{it}$ and the first-differenced regressors $\Delta \underline{x}_{jt}$ are stationary, the dependence between them at very distant leads and lags becomes trivial. We proceed by ignoring the truncation errors in the estimation. In Section 2, we show that doing so is asymptotically justified under the regularity conditions of Saikkonen (1991).[3]

There are three points worth noting here. First, including leads and lags of the first-differenced regressors $\underline{z}'_{it}$ in the cointegrating regressions controls for endogeneity but generally does not remove serial correlation. Therefore, in most applications it is likely that $u_{it}$ will be serially correlated. Second, $u^{\dagger}_{it}$ will probably be correlated with both leads and lags of the first-differenced regressors so it is necessary to include both leads and lags in the estimation.[4] Third, in the system environment, it is important to include leads and lags of the regressors from cross-equations in addition to own equation regressors. That is, the parametric adjustment for endogeneity in equation $i = 1$ will generally require including 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 $\Delta \underline{x}_{2t}$ to $\Delta \underline{x}_{Nt}$.

### *1.1. DSUR*

Substituting (3) into (1) and ignoring the truncation error yields the regression $y_{it} = \underline{x}'_{it}\underline{\beta}_i + \underline{z}'_{pt}\underline{\delta}_{pi} + u_{it}$. Let $\underline{y}_t = (y_{1t}, \ldots, y_{Nt})'$, $\underline{u}_t = (u_{1t}, \ldots, u_{Nt})'$, $\underline{\beta} = (\underline{\beta}'_1, \ldots, \underline{\beta}_N)'$, $\underline{\delta}_p = (\underline{\delta}'_{p1}, \ldots, \underline{\delta}'_{pN})'$, $\mathbf{Z}_{pt} = (\mathbf{I}_N \otimes \underline{z}_{pt})$, $\mathbf{X}_t = \text{diag}(\underline{x}_{1t}, \ldots, \underline{x}_{Nt})$ and $\mathbf{W}_t = (\mathbf{X}'_t, \mathbf{Z}'_{pt})'$. Then the equations can be stacked together in a system as

$$\underline{y}_t = (\underline{\beta}', \underline{\delta}'_p)\mathbf{W}_t + \underline{u}_t. \tag{5}$$

3. For ease of notation we assume an equal number $p$ of leads and lags but the extension of the analysis to asymmetric numbers of leads and lags is straightforward. Stock and Watson (1993), for example, assume that the equilibrium errors are correlated only with a finite number of leads and lags of the first-differenced regressors.

4. In investigations of predictions of rational expectations models, one might think that lags are unnecessary because the disturbance term has zero expectation conditional on the information available to the agents. Alternatively, it might be reasoned that leads are unnecessary because lags will render the disturbance term serially uncorrelated. As can be seen from our analysis, neither of these arguments is correct.

Denote the long-run covariance matrix of $\underline{u}_t$ by $\mathbf{\Omega}_{uu}$. The DSUR estimator with known $\mathbf{\Omega}_{uu}$ is

$$\begin{bmatrix} \hat{\underline{\beta}}_{\text{dsur}} \\ \hat{\underline{\delta}}_{p,\text{dsur}} \end{bmatrix} = \left(\sum_{t=p+1}^{T-p} \mathbf{W}_t \mathbf{\Omega}_{uu}^{-1} \mathbf{W}_t'\right)^{-1} \left(\sum_{t=p+1}^{T-p} \mathbf{W}_t \mathbf{\Omega}_{uu}^{-1} \underline{y}_t\right). \tag{6}$$

We show in Section 2.1.1 that $\hat{\underline{\beta}}_{\text{dsur}}$ is asymptotically mixed normal. It follows that tests of the linear restrictions $\mathbf{R}\underline{\beta} = \underline{r}$, where $\mathbf{R}$ is a $q \times Nk$ matrix of constants and $\underline{r}$ is a $q$-dimensional vector of constants, can be constructed by letting $\hat{\mathbf{V}}_{\text{dsur}} = \sum_{t=p+1}^{T-p} \mathbf{X}_t \mathbf{\Omega}_{uu}^{-1} \mathbf{X}_t'$ and forming the Wald statistic $W_{\text{dsur}} = \left(\mathbf{R}\hat{\underline{\beta}}_{\text{dsur}} - \underline{r}\right)'[\mathbf{R}\hat{\mathbf{V}}_{\text{dsur}}\mathbf{R}']^{-1}\left(\mathbf{R}\hat{\underline{\beta}}_{\text{dsur}} - \underline{r}\right)$ which is asymptotically distributed as a chi-square variate with $q$ degrees of freedom under the null hypothesis. This Wald statistic provides a convenient test of homogeneity restrictions on the cointegrating vectors, $H_0 : \beta_1 = \cdots = \beta_N$.

In applications, we replace $\mathbf{\Omega}_{uu}$ with a consistent estimator, $\hat{\mathbf{\Omega}}_{uu} \xrightarrow{p} \mathbf{\Omega}_{uu}$. Estimation of the long-run covariance matrix is discussed below. Such an estimator might be called a "feasible" DSUR estimator. It is straightforward to see that the asymptotic distribution of the feasible DSUR estimator is identical to that of the DSUR estimator where $\mathbf{\Omega}_{uu}$ is known. Accordingly, we will not make a distinction between estimators formed with a known $\mathbf{\Omega}_{uu}$ or one that is estimated.

Because the parametric control for endogeneity takes up degrees of freedom, DSUR is applicable where $N$ is substantially smaller than $T$. For $T = 100$, we show below in our Monte Carlo experiments and in our applications that systems with $N = 8$ are feasibly estimated.

### 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. When the numbers of included leads and lags $p$ 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. Under standard regularity conditions, the two-step DSUR estimator is asymptotically equivalent to the DSUR estimator $\hat{\underline{\beta}}_{\text{dsur}}$ discussed above.

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

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

and $\hat{\underline{\delta}}_{pi,\text{ols}} = \hat{\underline{\gamma}}_{pi}^{y} - \underline{\beta}_i'\hat{\underline{\gamma}}_{pi}^{x}$. Stacking the equations together as $\hat{\underline{y}}_t = \hat{\mathbf{X}}_t'\underline{\beta} + \hat{\underline{u}}_t$ and running SUR gives the two-step DSUR estimator,

$$\hat{\underline{\beta}}_{2\text{sdsur}} = \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{X}}_t \mathbf{\Omega}_{uu}^{-1} \hat{\mathbf{X}}_t'\right]^{-1} \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{X}}_t \mathbf{\Omega}_{uu}^{-1} \hat{\underline{y}}_t\right]. \tag{7}$$

### 1.3. Restricted DSUR

We now turn to estimation of the cointegration vector under the homogeneity restrictions $\underline{\beta}_1 = \cdots = \underline{\beta}_N = \underline{\beta}$. 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 now becomes one of estimating $\underline{\beta}$, in the system of equations $\hat{y}_{it} = \hat{\underline{x}}'_{it}\underline{\beta} + \hat{u}_{it}$. Stacking the equations together gives

$$\hat{\underline{y}}_t = \hat{\mathbf{x}}'_t \underline{\beta} + \hat{\underline{u}}_t, \tag{8}$$

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

Let $\mathbf{\Omega}_{uu} = \mathbf{LL}'$ be the lower triangular Choleski decomposition of the long-run error covariance matrix. Premultiply (8) by $\mathbf{L}^{-1}$ to get $\hat{\underline{y}}^*_t = \hat{\mathbf{x}}^{*'}_t \underline{\beta} + \hat{\underline{u}}^*_t$ 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:

$$\begin{aligned}\hat{\underline{\beta}}_{\text{rdsur}} &= \left[\sum_{i=1}^{N}\sum_{t=p+1}^{T-p} \hat{\underline{x}}^*_{it}\hat{\underline{x}}^{*'}_{it}\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=p+1}^{T-p} \hat{\underline{x}}^*_{it}\hat{\underline{y}}^*_{it}\right] \\ &= \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{x}}_t \mathbf{\Omega}_{uu}^{-1}\hat{\mathbf{x}}'_t\right]^{-1}\left[\sum_{t=p+1}^{T-p} \hat{\mathbf{x}}_t \mathbf{\Omega}_{uu}^{-1}\hat{\underline{y}}_t\right].\end{aligned} \tag{9}$$

This estimator is also asymptotically mixed normal. Tests of the set of $q$ linear restrictions $\mathbf{R}\underline{\beta} = \underline{r}$ can be conducted by comparing the Wald statistic $W_{\text{rdsur}} = \left(\mathbf{R}\hat{\underline{\beta}}_{\text{rdsur}} - \underline{r}\right)'[\mathbf{R}\hat{\mathbf{V}}_{\text{rdsur}}\mathbf{R}']^{-1}\left(\mathbf{R}\hat{\underline{\beta}}_{\text{rdsur}} - \underline{r}\right)$ to the chi-square distribution with $q$ degrees of freedom where $\mathbf{R}$ is a $q \times k$ matrix of constants, $\underline{r}$ is a $q$-dimensional vector of constants and $\hat{\mathbf{V}}_{\text{rdsur}} = \sum_{t=p+1}^{T-p} \mathbf{x}_t \mathbf{\Omega}_{uu}^{-1}\mathbf{x}'_t$.

### 1.4. Estimating the long-run covariance matrix

While many estimators of the long-run covariance matrix $\mathbf{\Omega}_{uu}$ are available, accurate estimation can become a challenge as $N$ increases. A single factor structure is an efficient parameterization that has been found to adequately model the long-run cross-sectional covariance.[5] We adopt the approach of Phillips and Sul (2003), which begins by assuming the parametric structure

$$u^{\dagger}_{it} = \underline{z}'_{pt}\underline{\delta}_{pi} + u_{it}, \tag{10}$$

$$u_{it} = \rho_i u_{it-1} + \sum_{j=1}^{n-1} \eta_{ij}\Delta u_{it-j} + w_{it}, \tag{11}$$

$$w_{it} = \lambda_i \theta_t + \zeta_{it}. \tag{12}$$

Equation (10) is a restatement of (3) that ignores the truncation error. In (11), potential serial correlation in $u_{it}$ is allowed for by letting it follow an AR($n$) process. The cross-sectional dependence is modelled by (12) where $\theta_t \overset{\text{i.i.d.}}{\sim} (0, 1)$ is the single factor, $\lambda_i$ is the factor loading for individual $i$ and $\zeta_{it}$ is an i.i.d. idiosyncratic error. Letting $\underline{w}_t = (w_{1t}, \ldots, w_{Nt})'$, $\underline{\zeta}_t = (\zeta_{1t}, \ldots, \zeta_{Nt})'$, $\underline{\lambda} = (\lambda_{1t}, \ldots, \lambda_{Nt})'$, $\mathbf{V}_{\zeta\zeta} = E\left(\underline{\zeta}_t\underline{\zeta}'_t\right)$, we have $\mathbf{V}_{ww} = E(\underline{w}_t\underline{w}'_t) = \mathbf{V}_{\zeta\zeta} + \underline{\lambda}\underline{\lambda}'$. Thus to estimate $\mathbf{\Omega}_{uu}$, proceed as follows:

**Step 1.** Run OLS on $y_{it} = a_i + \underline{x}'_{it}\underline{\beta}_i + \underline{z}'_{pt}\underline{\delta}_{pi} + u_{it}$ and obtain the quasi-residuals $\widehat{u}^*_{it} = \widehat{u}_{it} + \widehat{a}_i = y_{it} - \underline{x}'_{it}\widehat{\underline{\beta}}_i - \underline{z}'_{pt}\widehat{\underline{\delta}}_{pi}$.

5. See Bai and Ng (2002), Phillips and Sul (2003) and Moon and Perron (2004).

**Step 2.** Estimate the AR($n$) coefficients. To reduce bias, use recursive mean adjusted least squares by running OLS on $(\widehat{u}^*_{it} - \mu_{it-1}) = \rho_i(\widehat{u}^*_{it-1} - \mu_{it-1}) + \sum_{j=1}^{n-1} \eta_{ij}\, \Delta u^*_{it-j} + \xi_{it}$, where $\mu_{it-1} = (t-1)^{-1} \sum_{s=1}^{t-1} \widehat{u}^*_{is}$ is the recursive mean of the quasi-residuals. Using the estimated AR($n$) coefficients, recover the residuals, $\widehat{w}_{it} = \widehat{u}^*_{it} - \widehat{\rho}_i \widehat{u}^*_{it-1} - \sum_{j=1}^{n-1} \widehat{\eta}_{ij}\, \Delta\widehat{u}^*_{it-j}$.

**Step 3.** Form the sample error covariance matrix $\mathbf{M}_T = T^{-1} \sum_{t=1}^{T} \underline{\widehat{w}}_t \underline{\widehat{w}}'_t$. Estimate $\lambda_i$ and $\mathbf{V}_{\zeta\zeta}$ by the iterative method of moments, $(\widehat{\underline{\lambda}}, \widehat{\mathbf{V}}_{\zeta\zeta}) = \arg\min_{\underline{\lambda}, \mathbf{V}_{\zeta\zeta}} \{\mathrm{tr}[(\mathbf{M}_T - \mathbf{V}_{\zeta\zeta} - \underline{\lambda}\underline{\lambda}')(\mathbf{M}_T - \mathbf{V}_{\zeta\zeta} - \underline{\lambda}\underline{\lambda}')']\}$. Use the estimates to form the parametric estimate $\widehat{\mathbf{V}}_{ww} = \widehat{\mathbf{V}}_{\zeta\zeta} + \widehat{\underline{\lambda}}\,\widehat{\underline{\lambda}}'$.

**Step 4.** Obtain the estimate of the long-run covariance matrix, $\widehat{\boldsymbol{\Omega}}_{uu} = (\mathbf{I} - \underline{\widehat{\rho}})^{-1} \widehat{\mathbf{V}}_{ww} (\mathbf{I} - \underline{\widehat{\rho}})^{-1}$, where $\underline{\widehat{\rho}} = (\widehat{\rho}_{1t}, \ldots, \widehat{\rho}_{Nt})'$.

### *1.5. Lead and lag length selection*

An important problem in practice concerns the choice of $p$. 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 $(p^+_{ij}, p^-_{ij}) = (p^+_{ii}, p^-_{ii})$.

## 2. SAMPLING PROPERTIES OF DSUR AND SOME ALTERNATIVE COINTEGRATION VECTOR ESTIMATORS

Section 2.1 discusses the asymptotic properties of DSUR in comparison to alternative parametric cointegration vector estimators. The Monte Carlo experiments are discussed in Section 2.2.

### *2.1. Asymptotic properties*

Let $\underline{W}(r)$ be a vector standard Brownian motion for $0 \le r \le 1$, and let $[Tr]$ denote the largest integer value of $Tr$ for $0 \le r \le 1$. 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 ones such as $\int_0^1 \underline{W}(r)d\underline{W}(r)'$ are written as $\int \underline{W}d\underline{W}'$. Scaled vector Brownian motions are denoted by $\underline{B} = \boldsymbol{\Lambda}\underline{W}$ where $\boldsymbol{\Lambda}$ is a scaling matrix.

It follows from Assumption 1 that $\underline{w}^\dagger_t$ obeys the functional central limit theorem, $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \underline{w}^\dagger_t \overset{D}{\to} \underline{B}^\dagger(r) = \boldsymbol{\Psi}^\dagger(1)\underline{W}(r)$ where $\underline{B}^\dagger = (\underline{B}^\dagger_u, \underline{B}'_{e_1}, \ldots, \underline{B}'_{e_N})'$ is an $N(k+1)$-dimensional scaled vector Brownian motion with covariance matrix, $\boldsymbol{\Omega}^\dagger = \boldsymbol{\Psi}^\dagger(1)\boldsymbol{\Psi}^\dagger(1)' = \sum_{j=-\infty}^{\infty} \mathrm{E}[\underline{w}^\dagger_j \underline{w}^{\dagger'}_0] = \boldsymbol{\Gamma}^\dagger_0 + \sum_{j=1}^{\infty}(\boldsymbol{\Gamma}^\dagger_j + \boldsymbol{\Gamma}^{\dagger'}_j)$. The long-run covariance matrix and its components can be partitioned as

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

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

where $\mathbf{\Omega}_{e_iu}^\dagger$ is the long-run covariance between $\underline{e}_{it}$ and $(u_{1t}^\dagger, \ldots, u_{Nt}^\dagger)$, $i = 1, \ldots, N$, $\mathbf{\Gamma}_{uu,j}^\dagger = \mathrm{E}(\underline{u}_t^\dagger \underline{u}_{t-j}^{\dagger\prime})$, $\mathbf{\Gamma}_{ue_k,j}^\dagger = \mathrm{E}(\underline{u}_t^\dagger \underline{e}_{kt-j}^\prime)$, and $\mathbf{\Gamma}_{e_ke_s,j}^\dagger = \mathrm{E}(\underline{e}_{kt}\underline{e}_{st-j}^\prime)$.

We reintroduce the truncation errors by letting $\underline{v}_{pt} = (v_{p1t}, \ldots, v_{pNt})'$ and rewriting (5) as $\underline{y}_t = (\underline{\beta}', \underline{\delta}_p')\mathbf{W}_t + \underline{v}_{pt} + \underline{u}_t$. We ensure that the truncation errors vanish asymptotically by following Saikkonen (1991) with:

**Assumption 2 (Lead and Lag Dependence).** *Let $p(T)$ be the number of leads and lags of $\Delta\underline{x}_{it}$ $(i = 1, \ldots, N)$, included in* (3)*. We assume that*

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

(ii) $\sqrt{T}\sum_{|j|>p(T)} \left\| \begin{matrix} \underline{\delta}_{11,j}' & \cdots & \underline{\delta}_{1N,j}' \\ \vdots & \ddots & \vdots \\ \underline{\delta}_{N1,j}' & \cdots & \underline{\delta}_{NN,j}' \end{matrix} \right\| \to 0,$

*where $\|\cdot\|$ is the Euclidean norm.*

The second condition in Assumption 2 places an upper bound on the allowable dependence of $u_{it}^\dagger$ 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 specification error to vanish.

### 2.1.1. DSUR asymptotic properties.

We are now in a position to state[6]

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

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

(ii) *If* $\mathbf{B}_e = \mathrm{diag}(\underline{B}_{e_1}, \ldots, \underline{B}_{e_N})$, $\hat{\mathbf{V}}_{\mathrm{dsur}} = \sum_{t=p+1}^{T-p} \mathbf{X}_t \mathbf{\Omega}_{uu}^{-1}\mathbf{X}_t'$, *and* $\mathbf{R}$ *is a* $q \times Nk$ *matrix of constants such that* $\mathbf{R}\underline{\beta} = \underline{r}$, *then as* $T_* \to \infty$,

$$T_*(\hat{\underline{\beta}}_{\mathrm{dsur}} - \underline{\beta}) \xrightarrow{D} \left(\int \mathbf{B}_e \mathbf{\Omega}_{uu}^{-1}\mathbf{B}_e'\right)^{-1} \int \mathbf{B}_e \mathbf{\Omega}_{uu}^{-1} d\underline{B}_u, \tag{13}$$

*and*

$$(\mathbf{R}\hat{\underline{\beta}}_{\mathrm{dsur}} - \underline{r})'[\mathbf{R}\hat{\mathbf{V}}_{\mathrm{dsur}}\mathbf{R}']^{-1}(\mathbf{R}\hat{\underline{\beta}}_{\mathrm{dsur}} - \underline{r}) \xrightarrow{D} \chi_q^2. \tag{14}$$

The intuition behind Proposition 1 is that asymptotically, as the effects of the truncation error become trivial, one obtains a newly defined vector process $\underline{w}_t' = (u_{1t}, \ldots, u_{Nt}, \underline{e}_{1t}', \ldots, \underline{e}_{Nt}')$, with the moving average representation $\underline{w}_t = \mathrm{diag}[\mathbf{\Psi}_{11}(L), \mathbf{\Psi}_{22}(L)](\underline{\varepsilon}_{1t}', \underline{\varepsilon}_{2t}')'$, where

6. We follow Saikkonen (1991) in adopting a "degrees of freedom" adjustment, even though the asymptotic results are obviously the same without this adjustment.

$\mathbf{\Psi}_{11}(L)$ and $\mathbf{\Psi}_{22}(L)$ are $(N \times N)$ and $(Nk \times Nk)$ matrix polynomials in the lag operator $L$, respectively. It 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 the long-run covariance matrix $\mathbf{\Omega} = \text{diag}[\mathbf{\Omega}_{uu}, \mathbf{\Omega}_{ee}]$. Due to the block diagonality of $\mathbf{\Omega}$, it can be seen that $\underline{B}_u$ and $\underline{B}_e$ are independent. It is straightforward to see that the asymptotic distribution of the feasible DSUR estimator is identical to that of the DSUR estimator of Proposition 1.

As discussed above, some computational convenience is achieved by the two-step procedure. This entails no sacrifice in terms of asymptotic efficiency as seen in:

**Proposition 2 (Asymptotic Equivalence of the Two-Step Estimator).** *Under the conditions of Assumptions* 1 *and* 2*, the two-step DSUR estimator* (7) *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 in the first step and* $\underline{\beta}$ *estimated by running SUR on the GLS residuals in the second step.*

The asymptotic equivalence obtains due to the consistency of $\hat{\underline{\delta}}_{pi,\text{ols}}$ 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.

Under cointegration vector homogeneity, $\underline{\beta}_1 = \cdots = \underline{\beta}_N = \underline{\beta}$, we must restrict the amount of long-run dependence among regressors across equations. These restrictions are given in:

**Assumption 3.** $\mathbf{\Omega}_{ee}^{\dagger}$ *has full rank.*

Assumption 3 does not permit a common regressor across equations, nor does it allow the regressors across equations to be cointegrated. The properties of restricted DSUR are given in[7]

**Proposition 3 (Asymptotic Distribution of Restricted DSUR).** *Let* $\mathbf{b}_e = (\underline{B}_{e_1}, \ldots, \underline{B}_{e_N})$, $\mathbf{R}$ *be a* $q \times k$ *matrix of constants such that* $\mathbf{R}\hat{\underline{\beta}}_{\text{rdsur}} = \underline{r}$ *and* $\hat{\mathbf{V}}_{\text{rdsur}} = \sum_{t=p+1}^{T-p} \mathbf{x}_t \mathbf{\Omega}_{uu}^{-1} \mathbf{x}_t'$. *Then under the conditions of Assumptions* 1–3*, as* $T_* \to \infty$,

$$T_*(\hat{\underline{\beta}}_{\text{rdsur}} - \underline{\beta}) \xrightarrow{D} \left( \int \mathbf{b}_e \mathbf{\Omega}_{uu}^{-1} \mathbf{b}_e' \right)^{-1} \left( \int \mathbf{b}_e \mathbf{\Omega}_{uu}^{-1} d\underline{B}_u \right), \tag{15}$$

*and*

$$\left(\mathbf{R}\hat{\underline{\beta}}_{\text{rdsur}} - \underline{r}\right)' [\mathbf{R}\hat{\mathbf{V}}_{\text{rdsur}}\mathbf{R}']^{-1} \left(\mathbf{R}\hat{\underline{\beta}}_{\text{rdsur}} - \underline{r}\right) \xrightarrow{D} \chi_q^2. \tag{16}$$

**2.1.2. Alternative parametric estimators.** We now discuss the asymptotic properties of DOLS, system DOLS, and SSE for estimating heterogeneous cointegration vectors and panel DOLS for estimating homogeneous cointegration vectors.

7. In matrix notation, let $\hat{\underline{Y}}_T = (\hat{\underline{Y}}_1', \ldots, \hat{\underline{Y}}_N')'$ where $\hat{\underline{Y}}_i = (\hat{\underline{y}}_{ip+1}, \ldots, \hat{\underline{y}}_{iT-p})'$, $\hat{\mathbf{X}}_T = (\hat{\mathbf{X}}_1', \ldots, \hat{\mathbf{X}}_N')'$, $\hat{\mathbf{X}}_i = (\hat{\underline{x}}_{ip+1}, \ldots, \hat{\underline{x}}_{iT-p})'$ is the $T_* \times k$ matrix of regressors, and $\hat{\underline{u}}_T = (\hat{\underline{u}}_1, \ldots, \hat{\underline{u}}_N)'$, $\hat{\underline{u}}_i = (\hat{u}_{ip+1}, \ldots, \hat{u}_{iT-p})'$. The stacked system of observations is $\hat{\underline{Y}}_T = \hat{\mathbf{X}}_T \underline{\beta} + \hat{\underline{u}}_T$ where $\hat{\underline{\beta}}_{\text{rdsur}} = [\hat{\mathbf{X}}_T' (\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{I}_T) \hat{\mathbf{X}}_T]^{-1} [\hat{\mathbf{X}}_T' (\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{I}_T) \hat{\underline{Y}}_T]$.

*DOLS and system DOLS.* DOLS ignores dependence across individuals in estimation. Controlling for endogeneity in equation $i$ can be achieved by projecting $u_{it}^{\dagger}$ onto $\underline{z}_{pit}$ or onto $\underline{z}_{pt} = (\underline{z}'_{p1t}, \ldots, \underline{z}'_{pNt})'$ 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 only on $\underline{z}_{pit}$ DOLS.

While the joint distribution of DOLS across equations depends on the long-run covariance matrix $\mathbf{\Omega}_{uu}$, 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 is purged by least squares and the cointegration vector is 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 $\hat{\underline{y}}_t = \hat{\mathbf{X}}'_t \underline{\beta} + \hat{\underline{u}}_t$, where the dimensionalities of the matrices are as defined above. The system DOLS estimator is[8]

$$\hat{\underline{\beta}}_{\text{sysdols}} = \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{X}}_t \hat{\mathbf{X}}'_t\right]^{-1} \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{X}}_t \hat{\underline{y}}_t\right], \tag{17}$$

for which we have:

**Proposition 4 (Asymptotic Distribution of System DOLS).** *Under the conditions of Assumptions* 1 *and* 2, *as* $T_* \to \infty$,

$$T_*\left(\hat{\underline{\beta}}_{\text{sysdols}} - \underline{\beta}\right) \xrightarrow{D} \left(\int \mathbf{B}_e \mathbf{B}'_e\right)^{-1} \left(\int \mathbf{B}_e d\underline{B}_u\right), \tag{18}$$

*and*

$$\left(\mathbf{R}\hat{\underline{\beta}}_{\text{sysdols}} - \underline{r}\right)' [\mathbf{R}\hat{\mathbf{V}}_{\text{sysdols}}\mathbf{R}']^{-1} \left(\mathbf{R}\hat{\underline{\beta}}_{\text{sysdols}} - \underline{r}\right) \xrightarrow{D} \chi_q^2, \tag{19}$$

*where* $\hat{\mathbf{V}}_{\text{sysdols}} = \left[\sum_{t=p+1}^{T-p} \mathbf{X}_t \mathbf{X}'_t\right]^{-1} \left[\sum_{t=p+1}^{T-p} \mathbf{X}_t \mathbf{\Omega}_{uu} \mathbf{X}'_t\right] \left[\sum_{t=p+1}^{T-p} \mathbf{X}_t \mathbf{X}'_t\right]^{-1}$ *and* $\mathbf{R}$ *is a* $q \times Nk$ *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 $y_{it} = \underline{x}'_{it}\underline{\beta}_i + \underline{z}'_{it}\underline{\delta}_i + \zeta_{it}$, where $\zeta_{it}$ is the projection error which is by construction orthogonal to included leads and lags of $\Delta\underline{x}_{it}$. Since $(1/\sqrt{T})\sum(\zeta_{it}, \underline{e}'_{it})' \xrightarrow{D} \left(B_{\zeta i}, \underline{B}'_{ei}\right)'$ with long-run covariance matrix $\text{diag}\left(\Omega_{\zeta_i,\zeta_i}, \mathbf{\Omega}_{e_i,e_i}\right)$, it follows that conditional

8. Let $\widehat{\mathbf{x}}_i = \begin{bmatrix} \hat{\underline{x}}'_{ip+1} \\ \vdots \\ \hat{\underline{x}}'_{iT-p} \end{bmatrix}$ and $\mathbf{X}_T = \text{diag}[\widehat{\mathbf{x}}_1, \ldots, \widehat{\mathbf{x}}_N]$. Then in the standard matrix notation, $\mathbf{V}_{\text{sysdols}} = (\mathbf{X}'_T\mathbf{X}_T)^{-1}\mathbf{X}'_T(\mathbf{\Omega}_{uu} \otimes \mathbf{I}_T)\mathbf{X}_T(\mathbf{X}'_T\mathbf{X}_T)^{-1}$.

on $\underline{B}_{e_i}$, $\text{avar}(\hat{\underline{\beta}}_{\text{dols}}) = \Omega_{\zeta_i,\zeta_i}\left(\int \underline{B}_{e_i}\underline{B}'_{e_i}\right)^{-1}$. Since $\Omega_{\zeta_i,\zeta_i}$ is the long-run variance of the error from projecting $u^{\dagger}_{it}$ onto $\underline{z}_{pit} \subseteq \underline{z}_{pt}$ and $\Omega_{u_i,u_i}$ is the long-run variance of the error from projecting $u^{\dagger}_{it}$ onto $\underline{z}_{pt}$, it must be the case that $\Omega_{\zeta_i,\zeta_i} \geq \Omega_{u_i,u_i}$. Thus, $\text{avar}(\hat{\underline{\beta}}_{i,\text{dols}}) \geq \text{avar}(\hat{\underline{\beta}}_{i,\text{sysdols}})$.

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

**Proposition 5.** *Under the conditions of Assumptions* 1 *and* 2,

$$\text{avar}(\hat{\underline{\beta}}_{\text{dsur}}) \leq \text{avar}(\hat{\underline{\beta}}_{\text{sysdols}}).$$

DSUR is asymptotically efficient relative to the DOLS methods when the equilibrium errors exhibit cross-sectional dependence and when regressors in the cointegrating regression are heterogeneous across individuals. The latter condition is typical in panel cointegration analysis. It is straightforward to see that system DOLS and DSUR are asymptotically equivalent if the off-diagonal elements of $\mathbf{\Omega}_{uu}$ are zero.

*Saikkonen's system estimator.* Saikkonen (1993) proposed an estimator for a structural system of cointegrating equations. For presentation of Saikkonen's estimator, let $\underline{y}_t = (y_{1t}, \ldots, y_{Nt})'$, $\underline{x}_t = (\underline{x}_{1t}, \ldots, \underline{x}_{nt})'$, and $\underline{u}_t = (u_{1t}, \ldots, u_{Nt})'$. Saikkonen considered the structural simultaneous equation system $\mathbf{B}\underline{y}_t = \mathbf{C}\underline{x}_t + \underline{e}_t$ where $\mathbf{B}$ and $\mathbf{C}$ are matrices of structural coefficients. When $\mathbf{B}$ is non-singular, premultiplication gives $\underline{y}_t = \mathbf{A}\underline{x}_t + \underline{u}_t$, which is the system of cointegrating regressions stacked together where $\mathbf{A} = \mathbf{B}^{-1}\mathbf{C}$, and $\underline{u}_t = \mathbf{B}^{-1}\underline{e}_t$. While DSUR is an estimator of the reduced form coefficients $\mathbf{A}$, Saikkonen's primary interest lies in estimation of the structural coefficients when linear identification restrictions are available. The coefficient vector of interest is $\underline{\beta}_i = (-\underline{\theta}'_i, \underline{\gamma}'_i)'$ where $\underline{\theta}_i$ and $\underline{\gamma}_i$ are column vectors formed by the non-zero elements in row $i$ of $\mathbf{B} - \mathbf{I}_N$ and $\mathbf{C}$, respectively. Now let $\mathbf{H}$ be the "selection matrix" that gives $\mathbf{H}'(\mathbf{\Omega}_{uu}^{-1} \otimes \sum_{t=1}^{T} \underline{x}'_t\underline{x}_t)\mathbf{H} = \sum_{t=1}^{T} \mathbf{X}_t\mathbf{\Omega}_{uu}^{-1}\mathbf{X}_t$.[9] Then Saikkonen's system estimator (SSE) is

$$\widehat{\underline{\beta}}_{\text{SSE}} = \left[\mathbf{H}'\left(\mathbf{\Omega}_{uu}^{-1} \otimes \sum\nolimits_{t=1}^{T} \underline{x}'_t\underline{x}_t\right)\mathbf{H}\right]^{-1} \mathbf{H}'\left(\mathbf{\Omega}_{uu}^{-1} \otimes \sum\nolimits_{t=1}^{T} \underline{x}'_t\underline{x}_t\right)\underline{\hat{a}}, \tag{20}$$

where $\widehat{\underline{a}} = vec(\widehat{\mathbf{A}}')$, $\widehat{\mathbf{A}}' = \left[\sum_{t=1}^{T} \widehat{\underline{x}}'_t\widehat{\underline{x}}_t\right]^{-1} \sum_{t=1}^{T} \widehat{\underline{x}}_t[\widehat{y}_{1t}, \ldots, \widehat{y}_{Nt}]$ is a system estimator of the reduced form coefficients and $\widehat{\underline{x}}_t$ is the vector of regression errors from regressing $\underline{x}_t$ on $\underline{z}_{pt}$, and similarly for $\widehat{y}_{it}$.

SSE and DSUR are only comparable when the reduced form and the structure are the same ($\mathbf{B} = \mathbf{I}_N$). In this case, SSE and DSUR are asymptotically equivalent. However, this equivalence does not hold in finite samples where DSUR is more efficient. It can be shown, albeit under the more restrictive regularity conditions of Ogaki and Choi (2001), that DSUR is the conditionally Best Linear Unbiased Estimator (BLUE).[10] Although SSE with a known long-run covariance matrix is both linear and unbiased, it is not identical to DSUR and is therefore less efficient for finite $T$. To gain intuition for why this is so, we draw on a parallel result in the case of a single cointegrating regression. There, dynamic generalized

9. See Saikkonen (1993) for construction of the selector matrix $\mathbf{H}$.

10. They assume that strict exogeneity can be achieved by including a fixed number of leads and lags of the first-differenced regressors and that the conditional error covariance matrix is known.

least squares (DGLS) is asymptotically equivalent to DOLS (Phillips and Park (1988), Jo and Park (1997)), but a similar Ogaki and Choi analysis finds that DGLS is conditionally BLUE. Below, we find in our Monte Carlo experiments that DSUR is substantially more efficient than SSE in small samples even when some of Ogaki and Choi's regularity conditions are violated.

We note also that the Moon (1999) fully modified SUR estimator is linear unbiased when the long-run covariance matrix is known. By the same reasoning, DSUR is more efficient under the Ogaki and Choi conditions. On the other hand, the Park and Ogaki (1991) non-parametric Seemingly Unrelated Canonical Cointegrating Regression estimator is not linear and cannot be directly compared. What we can say, however, is that if strict exogeneity can be achieved by parametric adjustment, then DSUR will dominate the non-parametric estimators, but if the parametric assumption is violated, it is possible that robust non-parametric estimators will dominate parametric approaches.

*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

$$\hat{\underline{\beta}}_{\text{pdols}} = \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{x}}_t \hat{\mathbf{x}}_t'\right]^{-1} \left[\sum_{t=p+1}^{T-p} \hat{\mathbf{x}}_t \hat{\underline{y}}_t\right], \tag{21}$$

where $\hat{\mathbf{x}}_t = (\hat{\underline{x}}_{1t}, \dots, \hat{\underline{x}}_{Nt})$ is a $k \times N$ matrix. The asymptotic sampling properties of panel DOLS under cross-sectional dependence are given as a corollary to Proposition 4.

**Corollary 6 (Asymptotic Distribution of Panel DOLS).** *Let* $\mathbf{b}_e = (\underline{B}_{e_1}, \dots, \underline{B}_{e_N})$, $\hat{\mathbf{V}}_{\text{pdols}} = \left[\sum_{t=p+1}^{T-p} \mathbf{x}_t \mathbf{x}_t'\right]^{-1} \left[\sum_{t=p+1}^{T-p} \mathbf{x}_t \mathbf{\Omega}_{uu} \mathbf{x}_t'\right] \left[\sum_{t=p+1}^{T-p} \mathbf{x}_t \mathbf{x}_t'\right]^{-1}$*, and* $\mathbf{R}$ *be a* $q \times Nk$ *matrix of constants such that* $\mathbf{R}\underline{\beta} = \underline{r}$*. Then under the conditions of Assumptions* 1–3*, as* $T_* \to \infty$,

$$T_*\left(\hat{\underline{\beta}}_{\text{pdols}} - \underline{\beta}\right) \xrightarrow{D} \left(\int \mathbf{b}_e \mathbf{b}_e'\right)^{-1} \int \mathbf{b}_e d\underline{B}_u, \tag{22}$$

*and*

$$\left(\mathbf{R}\hat{\underline{\beta}}_{\text{pdols}} - \underline{r}\right)' [\mathbf{R}\hat{\mathbf{V}}_{\text{pdols}}\mathbf{R}']^{-1} \left(\mathbf{R}\hat{\underline{\beta}}_{\text{pdols}} - \underline{r}\right) \xrightarrow{D} \chi_q^2. \tag{23}$$

Finally, it is obvious that $\text{avar}\left(\hat{\underline{\beta}}_{\text{rdsur}}\right) \le \text{avar}\left(\hat{\underline{\beta}}_{\text{pdols}}\right)$.

### 2.2. *Monte Carlo experiments*

In this section, we study the small sample properties of the estimators discussed above by way of a series of Monte Carlo experiments. First, we compare the performance of feasible DSUR, SSE, system DOLS, and DOLS in an environment where the cointegration vector exhibits heterogeneity across equations. Second, we compare feasible restricted DSUR, and panel DOLS in an environment where the cointegrating vector is identical across equations.

**2.2.1. Experimental design.** Each cointegrating regression has a single regressor. The general form of the data generating process (DGP) is given by

TABLE 1

*Monte Carlo performance of DSUR, system DOLS (SDOLS), and SSE (Saikkonen's system estimator) under cointegration vector heterogeneity (BIC: selection rule)*

| | True | DSUR percentiles | | | | | MSE relative to DOLS | | |
|---|---|---|---|---|---|---|---|---|---|
| $N$ | mean | $T$ | 5% | 50% | 95% | Mean | DSUR | SDOLS | SSE |
| | | Low cross-sectional dependence | | | | | | | |
| | | 100 | 1·359 | 1·524 | 1·712 | 1·528 | 0·781 | 0·861 | 4·535 |
| 3 | 1·506 | 200 | 1·437 | 1·514 | 1·604 | 1·517 | 0·646 | 0·844 | 3·504 |
| | | 300 | 1·462 | 1·512 | 1·570 | 1·513 | 0·609 | 0·831 | 2·907 |
| | | 100 | 1·273 | 1·446 | 1·630 | 1·448 | 0·568 | 0·818 | 7·819 |
| 5 | 1·437 | 200 | 1·362 | 1·441 | 1·526 | 1·442 | 0·467 | 0·777 | 6·318 |
| | | 300 | 1·388 | 1·440 | 1·497 | 1·441 | 0·445 | 0·773 | 6·188 |
| | | 100 | 1·326 | 1·483 | 1·647 | 1·484 | 0·430 | 0·764 | 13·862 |
| 8 | 1·478 | 200 | 1·409 | 1·480 | 1·555 | 1·481 | 0·340 | 0·706 | 11·166 |
| | | 300 | 1·434 | 1·479 | 1·528 | 1·480 | 0·322 | 0·693 | 10·280 |
| | | High cross-sectional dependence | | | | | | | |
| | | 100 | 1·279 | 1·512 | 1·745 | 1·512 | 0·379 | 1·003 | 6·385 |
| 3 | 1·506 | 200 | 1·405 | 1·508 | 1·609 | 1·508 | 0·295 | 0·970 | 6·012 |
| | | 300 | 1·442 | 1·507 | 1·572 | 1·507 | 0·251 | 0·965 | 5·927 |
| | | 100 | 1·224 | 1·442 | 1·658 | 1·442 | 0·261 | 1·018 | 12·442 |
| 5 | 1·437 | 200 | 1·346 | 1·439 | 1·533 | 1·439 | 0·197 | 0·964 | 12·661 |
| | | 300 | 1·380 | 1·438 | 1·498 | 1·438 | 0·163 | 0·946 | 13·112 |
| | | 100 | 1·304 | 1·481 | 1·663 | 1·482 | 0·201 | 1·010 | 23·721 |
| 8 | 1·478 | 200 | 1·402 | 1·479 | 1·557 | 1·479 | 0·150 | 0·924 | 24·580 |
| | | 300 | 1·430 | 1·478 | 1·529 | 1·479 | 0·132 | 0·917 | 25·650 |

$$
\begin{aligned}
y_{it} &= \beta_i x_{it} + u_{it}^{\dagger}, \\
u_{it}^{\dagger} &= \rho_i u_{it-1}^{\dagger} + \sum_{j=1}^{n-1} \delta_{ij} \Delta x_{it-1} + w_{it}, \\
\Delta x_{it} &= \phi_i \Delta x_{it-1} + m_{it}, \\
w_{it} &= \lambda_i \theta_t + \zeta_{it}.
\end{aligned}
$$

Cross-sectional dependence and cross-sectional endogeneity are modulated by varying $\rho_i$ and $\lambda_i$. For each experiment, we generate 5000 random samples of $T = 100, 200$, and 300 observations. The parameters are set according to $\rho_i \sim U[0{\cdot}2, 0{\cdot}8]$, $\phi_i \sim U[0{\cdot}1, 0{\cdot}2]$, $\delta_{ij} \sim U[0{\cdot}2, 0{\cdot}8]$ where $U[\cdot]$ is the uniform distribution. Under heterogeneous cointegration, we set $\beta_i \sim U[1, 2]$ whereas under homogeneous cointegration we set $\beta_i = 1$. Cross-sectional dependence is said to be low when $\lambda_i \sim U[0, 1]$ (correlation ranges between 0 and 0·5) and is said to be high when $\lambda_i \sim U[1, 3]$ (correlation ranges between 0·5 and 0·99).[11]

**2.2.2. Results.** Table 1 reports 5, 50, and 95 percentiles and the mean of the Monte Carlo distribution for DSUR, system DOLS, and SSE as well as their relative (to DOLS) mean square error. The relative efficiency of DSUR is seen to improve for a given $N$ as $T$ increases but also for given $T$ and as $N$ grows over the range of $N$ and $T$ that we consider. For $T = 100$, under low cross-sectional dependence, DSUR achieves a 22% reduction in mean square error relative to DOLS when $N = 3$ but achieves a 57% reduction when $N = 8$. Even greater relative efficiency gains are achieved when there is a high degree of cross-sectional dependence. System

11. We have experimented extensively with a variety of DGPs. In all the cases that we studied, we found that DSUR exhibits moderate to strong efficiency advantages over the comparison estimators. To economize on space, we report a small set of illustrative results. Results from alternative DGPs are available from the authors upon request.

TABLE 2

*Monte Carlo performance: DSUR tests of homogeneity restrictions*

| | | $H_0 : \beta_1 = \cdots = \beta_N$ | | | $H_0 : \beta_1 = \cdots = \beta_N = 1$ | | |
|---|---|---|---|---|---|---|---|
| $N$ | $T$ | 10% | 5% | 5% size | 10% | 5% | 5% size |
| | | Low cross-sectional dependence | | | | | |
| 3 | 100 | 17·676 | 8·694 | 0·104 | 26·794 | 13·662 | 0·141 |
| | 200 | 11·473 | 6·952 | 0·068 | 15·087 | 9·648 | 0·086 |
| | 300 | 9·563 | 5·875 | 0·047 | 12·612 | 8·135 | 0·056 |
| 5 | 100 | 31·261 | 19·166 | 0·234 | 39·531 | 23·412 | 0·274 |
| | 200 | 20·458 | 13·241 | 0·127 | 24·222 | 15·366 | 0·145 |
| | 300 | 17·294 | 11·258 | 0·092 | 20·151 | 13·725 | 0·104 |
| 8 | 100 | 54·803 | 33·78 | 0·395 | 58·203 | 39·126 | 0·430 |
| | 200 | 29·335 | 21·165 | 0·185 | 32·443 | 23·945 | 0·204 |
| | 300 | 26·107 | 19·595 | 0·150 | 28·39 | 21·35 | 0·162 |
| | | High cross-sectional dependence | | | | | |
| 3 | 100 | 14·62 | 8·378 | 0·106 | 20·4 | 11·821 | 0·131 |
| | 200 | 11·503 | 6·968 | 0·073 | 14·469 | 9·416 | 0·087 |
| | 300 | 10·672 | 6·674 | 0·063 | 13·672 | 8·824 | 0·071 |
| 5 | 100 | 21·809 | 13·532 | 0·125 | 25·255 | 16·706 | 0·145 |
| | 200 | 15·754 | 10·292 | 0·064 | 18·145 | 12·141 | 0·068 |
| | 300 | 14·517 | 9·603 | 0·052 | 16·567 | 11·285 | 0·053 |
| 8 | 100 | 34·39 | 23·113 | 0·207 | 38·695 | 25·861 | 0·224 |
| | 200 | 22·292 | 16·821 | 0·101 | 24·848 | 18·474 | 0·103 |
| | 300 | 22·004 | 16·104 | 0·087 | 24·23 | 17·882 | 0·087 |

DOLS is generally more efficient than DOLS. The relative performance of the SSE estimator is unsatisfactory. We conclude from Table 1 that DSUR offers substantial efficiency gains over DOLS, SDOLS, and SSE, especially when there is a high degree of cross-sectional dependence in the equilibrium errors.

We now turn to the small sample properties of Wald test statistics for DSUR tests of homogeneity in the cointegrating regression slope coefficient. The first test considered is $H_0 : \beta_1 = \cdots = \beta_N$, where the Wald statistic is asymptotically $\chi^2$ with $N - 1$ degrees of freedom under the null hypothesis. The second test is for $H_0 : \beta_1 = \cdots = \beta_N = 1$ and the Wald statistic here is asymptotically $\chi^2$ with $N$ degrees of freedom. Table 2 reports the percentiles of the empirical test statistic distribution that lies to the right of the asymptotic 10% and 5% critical values as well as the effective (5%) size of the tests. There is some size distortion in the tests which worsens somewhat as $N$ increases for given $T$. The tests are better sized under high cross-sectional dependence, which is consistent with Table 1 results that show better relative efficiency of DSUR under high cross-sectional dependence. The size distortion in the test of equality of coefficients is less severe than the test that all slope coefficients equal 1. The size distortions in both tests are quite moderate when $T = 300$.

The small sample performance of restricted DSUR for estimation under homogeneity constraints is reported in Table 3. In addition to selected percentiles of the distribution, the table displays the mean square error of restricted DSUR relative to panel DOLS. Here, it can be seen that restricted DSUR achieves sizable efficiency gains over panel DOLS. The relative performance of restricted DSUR is better under high cross-sectional dependence and generally improves as $N$ increases for fixed $T$.

We conclude that in small samples, efficiency gains are available for the DSUR methods, especially when there is moderate to strong cross-sectional dependence. For $T = 100$, the tests of

TABLE 3

*Monte Carlo performance of restricted DSUR (RDSUR)*

| | | RDSUR percentiles | | | | Relative |
|---|---|---|---|---|---|---|
| *N* | *T* | 5% | 50% | 95% | Mean | MSE |
| | | Low cross-sectional dependence | | | | |
| | 100 | 0·930 | 1·026 | 1·146 | 1·030 | 0·813 |
| 3 | 200 | 0·969 | 1·013 | 1·070 | 1·015 | 0·663 |
| | 300 | 0·980 | 1·008 | 1·044 | 1·010 | 0·616 |
| | 100 | 0·952 | 1·007 | 1·071 | 1·009 | 0·521 |
| 5 | 200 | 0·978 | 1·003 | 1·032 | 1·004 | 0·402 |
| | 300 | 0·986 | 1·002 | 1·021 | 1·003 | 0·382 |
| | 100 | 0·964 | 1·006 | 1·051 | 1·006 | 0·330 |
| 8 | 200 | 0·984 | 1·002 | 1·022 | 1·003 | 0·234 |
| | 300 | 0·990 | 1·001 | 1·014 | 1·001 | 0·223 |
| | | High cross-sectional dependence | | | | |
| | 100 | 0·904 | 1·000 | 1·098 | 1·000 | 0·205 |
| 3 | 200 | 0·956 | 0·999 | 1·042 | 0·999 | 0·115 |
| | 300 | 0·972 | 0·999 | 1·025 | 0·999 | 0·103 |
| | 100 | 0·942 | 1·001 | 1·060 | 1·001 | 0·181 |
| 5 | 200 | 0·974 | 1·000 | 1·028 | 1·000 | 0·124 |
| | 300 | 0·983 | 1·001 | 1·018 | 1·000 | 0·115 |
| | 100 | 0·957 | 1·001 | 1·047 | 1·002 | 0·274 |
| 8 | 200 | 0·981 | 1·000 | 1·019 | 1·000 | 0·109 |
| | 300 | 0·988 | 1·000 | 1·012 | 1·000 | 0·099 |

*Note*: MSE is the mean square error relative to panel DOLS. The true mean is 1.

homogeneity restrictions are somewhat oversized and use of the asymptotic theory in applications may lead to over-rejections of the null hypothesis. However, for $T = 300$, the DSUR tests are accurately 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 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. Our second application revisits the anomaly reported by Evans and Lewis (1995) that the expected excess return from forward foreign exchange rate speculation is unit-root non-stationary.

### *3.1. 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

$$\left(\frac{I}{Y}\right)_i = \alpha + \beta \left(\frac{S}{Y}\right)_i + u_i, \tag{24}$$

to test the hypothesis that capital is perfectly mobile internationally. They find that $\beta$ 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, Kulasi and Smith (1998)).

Theoretical studies, on the other hand, have shown that the Feldstein and Horioka (1980) logic is not airtight. Obstfeld (1986), Cantor and Mark (1988), Cole and Obstfeld (1991), and 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 of 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 non-stationary, they are cointegrated with a cointegrating vector $(1, -1)$. Thus the long-run relationship between saving and investment studied by means of 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, Jansen (1996), Coakley and Kulasi (1997), and Hussein (1998) show that the saving and investment ratios are cointegrated. Under this interpretation, the current account is a key component of the equilibrium error. Cross-sectional dependence arises naturally because the current accounts for all countries in the world must sum to zero.

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. In contrast to previous analyses which have employed non-system methods, we provide a direct test of the solvency constraint using an efficient system estimation technique that explicitly accounts for cross-country dependence. This is a natural application for DSUR as we seek efficient estimation for panel cointegration regression with a moderate cross-sectional component.

Since our focus is on the long-run relationship between saving and investment, we follow the Coakley *et al*. interpretation that the long-run solvency constraint implies cointegration. Two versions of their model imply slightly different forms of cointegration. First, if we assume that saving and investment are unit-root non-stationary, then this version of their model implies that the current account is stationary and saving and investment are cointegrated with cointegrating vector $(1, -1)$. Second, if we assume that the saving–GDP ratio and investment–GDP ratio are unit-root non-stationary, 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,

$$\left(\frac{I}{Y}\right)_{it} = \alpha_i + \beta_i \left(\frac{S}{Y}\right)_{it} + u_{it}. \tag{25}$$

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 balanced growth in our data-set.[12]

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

$$\ln(I_{it}) = \alpha_i + \beta_i \ln(S_{it}) + u_{it}. \tag{26}$$

With the relatively short time series available, it was not feasible to simultaneously estimate all 12 countries' regressions by DSUR due to the excessive number of parameters in the system. To proceed, we break the panel into subsamples and estimate separate systems for European and non-European countries.

Table 4 reports our estimates. We first discuss the results in ratio form. For the European countries, the BIC rule selects $p = 3$. Only our DSUR slope coefficient estimate for the U.K. is significantly different from 1. For non-European countries ($p = 3$), the point estimate is marginally significantly different from 1 only for the U.S. Tests of homogeneity provide little evidence against the hypothesis of slope coefficient equality. In the European system, the asymptotic $p$-value for the test of $H_0 : \beta_1 = \cdots = \beta_N$ is 0·31. The asymptotic $p$-value for the test of $H_0 : \beta_1 = \cdots = \beta_N = 1$ is 0·25. For the non-European system, neither of the tests for homogeneity can be rejected at the asymptotic 1% level. These results suggest that it is reasonable to pool and to re-estimate the two systems under homogeneity. When we do so, we obtain a restricted DSUR estimate 1·049 which is insignificantly different from 1 for the European system. The restricted DSUR estimate for the non-European system of 0·861 is also insignificantly different from 1.

Looking at the estimates from the log levels regression, the European data-set tells a similar story. These estimates, again associated with $p = 3$, are all insignificantly different from 1. Neither test of the homogeneity restrictions rejects at the 5% level. For the non-European countries, our BIC rule sets $p = 2$. Here, the DSUR estimate for the U.S. of 1·094 is significantly greater than 1. Since the homogeneity restrictions are not rejected, we re-estimate by restricted DSUR. This gives a point estimate of 0·989 which is insignificantly 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 the Coakley *et al*. solvency constraint is not violated.

12. We perform the Phillips and Sul (2003) panel unit-root test which is 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 and 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.

TABLE 4

*Saving–investment correlations*

| | Ratios | | Log levels | |
|---|---|---|---|---|
| | $\hat{\beta}_i$ | $t\ (\beta_i = 1)$ | $\hat{\beta}_i$ | $t\ (\beta_i = 1)$ |
| A. European system | | | | |
| Austria | 0·988 | −0·127 | 0·987 | −0·707 |
| Finland | 1·189 | 0·666 | 0·922 | −1·585 |
| France | 0·898 | −0·850 | 0·982 | −0·683 |
| Germany | 0·690 | −0·692 | 1·007 | 0·059 |
| Italy | 1·049 | 0·188 | 0·971 | −0·901 |
| Spain | 0·729 | −1·821 | 1·007 | 0·177 |
| Switzerland | 1·144 | 0·414 | 0·920 | −0·713 |
| U.K. | 0·566 | −2·398 | 1·002 | 0·058 |
| $\chi^2_7$ | 8·284 | | 2·513 | |
| ($p$-value) | (0·308) | | (0·926) | |
| $\chi^2_8$ | 10·917 | | 3·075 | |
| ($p$-value) | (0·251) | | (0·930) | |
| Restricted | 1·049 | 0·701 | 0·990 | −0·730 |
| B. Non-European system | | | | |
| Australia | 0·713 | −0·711 | 0·994 | −0·179 |
| Canada | 0·832 | −1·695 | 0·988 | −0·181 |
| Japan | 0·982 | −0·069 | 0·971 | −0·519 |
| U.S. | 0·803 | −1·632 | 1·094 | 2·367 |
| $\chi^2_3$ | 0·443 | | 3·666 | |
| ($p$-value) | (0·931) | | (0·300) | |
| $\chi^2_4$ | 6·996 | | 4·984 | |
| ($p$-value) | (0·136) | | (0·289) | |
| Restricted | 0·861 | −1·204 | 0·989 | −0·444 |

*Note*: The statistic for the test of homogeneity is $\chi^2_7$ in panel A and $\chi^2_3$ in panel B. The statistic for the test that slope coefficients are all equal to 1 is $\chi^2_8$ in panel A and $\chi^2_4$ in panel B.

### 3.2. *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 logarithm of the one-period forward exchange rate. It is widely agreed that since the move to generalized floating in 1973 both $s_{it} \sim I(1)$ and $f_{it} \sim I(1)$ and they are cointegrated. Let $\beta_i$ be the cointegrating coefficient of $s_{it+1}$ 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 non-stationary and is cointegrated with $f_{it}$. Evans and Lewis ask whether $p_{it}$ is $I(0)$ or $I(1)$, by estimating the cointegrating regression

$$s_{it+1} = \alpha_i + \beta_i f_{it} + u^{\dagger}_{it+1}. \tag{27}$$

They test the hypothesis $H_0 : \beta_i = 1$ using monthly observations from January 1975 to December 1989 on the dollar rates of the pound, deutschmark, and yen, and are able to reject that null hypothesis at small significance levels. The implied non-stationarity of the excess return is an anomaly.

TABLE 5

*DSUR estimation of spot and forward exchange rate cointegrating regression*, 1975.1–1996.12. *Three leads and lags*

| | $\hat{\beta}$ | $t(\beta = 1)$ |
|---|---|---|
| Germany | 0·992 | −0·024 |
| Japan | 0·998 | −0·800 |
| U.K. | 1·001 | 0·351 |
| $\chi_2^2$ | 0·539 | |
| ($p$-value) | (0·764) | |
| $\chi_3^2$ | 0·761 | |
| ($p$-value) | (0·859) | |
| Restricted | 0·998 | −1·833 |

*Notes*: $\chi_2^2$ is the test statistic for testing the homogeneity hypothesis $\beta_1 = \beta_2 = \beta_3$. $\chi_3^2$ is the test statistic for testing the homogeneity hypothesis $\beta_1 = \beta_2 = \beta_3 = 1$.

Evans and Lewis employ both DOLS and an SUR estimator, but in the latter, they included only the leads and lags of first-differenced regressors from the "own" equation and not from cross-equations. The omission of leads and lags of the first-differenced regressors from other equations can be problematic since this may not control for endogeneity in the dynamic regressions even asymptotically. Strict exogeneity would require that the error in each equation at a point in time must be uncorrelated with regressors from all equations in the system at any point in time. However, one component of the error is the forecast error of investors for the exchange rate and the examination is on the three exchange rates quoted relative to the U.S. dollar. If unexpected U.S. macroeconomic shocks have an effect on all three forward rates, the forecast error in each equation will be correlated with forward rates in the other two equations.

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, deutschmark, and yen relative to the U.S. dollar from January 1975 to December 1996. We have 286 time-series observations sampled from every fourth Friday of the Bank of Montreal/Harris Bank *Foreign Exchange Weekly Review*. The estimation results are reported in Table 5. In the light of the moderate size distortion uncovered in the Monte Carlo analysis, we test the hypotheses using the 1% 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 each of the currencies. We employ two tests of homogeneity in the cointegration vectors. The first one tests the null hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3$. The second is a test of the null hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 1$. Neither of these homogeneity restrictions are rejected at conventional significance levels. We proceed by imposing the homogeneity restrictions in estimation and obtain a restricted DSUR estimate that is insignificantly different from 1. We conclude that the evidence for non-stationarity of the excess return is less compelling according to the DSUR slope coefficient estimates under homogeneity restrictions.

## 4. CONCLUSION

In this paper, we proposed the DSUR estimator for multiple-equation cointegrating regressions for situations in which the cointegration vector displays heterogeneity across equations and in which it is homogeneous. This estimator exploits the cross-equation correlation in the equilibrium errors and it is efficient. DSUR estimators can be successfully applied in small to moderate systems where the number of time periods is substantially larger than the number of cointegrating equations. DSUR will not be feasible in systems of large $N$ due to the proliferation of free parameters that must be estimated in the error correlation.

In our application on the saving–investment regression, we used a natural classification of subsystems according to geography such that each subsample might reasonably exhibit high degrees of cross-equation equilibrium error correlation. Our empirical results were somewhat mixed, but the evidence suggested 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 the long-run solvency constraint is not violated.

For our application on the spot and forward exchange rates, we argued that it is problematic to omit leads and lags of the first-differenced regressors from other equations in applying the SUR method to dynamic cointegrating regressions as Evans and Lewis did. We conclude that the evidence for non-stationarity of the excess return is less compelling according to the DSUR slope coefficient estimates under homogeneity restrictions than originally found by Evans and Lewis.

## APPENDIX

Our asymptotic results are for $T \to \infty$ for fixed $N$. For notational convenience and without loss of generality, we take $N = 2$.

*Proof of Proposition* 1. We note that three regularity conditions assumed by Saikkonen (1991) are satisfied under Assumption 1. They are (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 fourth-order cumulants are absolutely summable. Let $T_* = T - 2p$,

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

and

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

Then

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

From Theorem 4.1 of Saikkonen (1991), we have $\mathbf{G}_T^{-1}\sum_{t=p+1}^{T-p}\mathbf{W}_t\boldsymbol{\Omega}_{uu}^{-1}\underline{v}_{pt} = o_p(1)$ 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 $T_*(\underline{\hat{\beta}}_{\text{dsur}} - \underline{\beta})$ and $\sqrt{T_*}(\underline{\hat{\delta}}_{p,\text{dsur}} - \underline{\delta}_p)$ are asymptotically independent. It follows that

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

Conditional on $\mathbf{B}_e$, $\int \mathbf{B}_e \boldsymbol{\Omega}_{uu}^{-1} d\underline{B}_u \xrightarrow{D} \text{N}(\underline{0}, [\int \mathbf{B}_e \boldsymbol{\Omega}_{uu}^{-1}\mathbf{B}_e'])$ (Park and Phillips, 1988). 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 $\mathbf{B}_e$,

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

Since the chi-square distribution does not depend on $\int \mathbf{B}_e \boldsymbol{\Omega}_{uu}^{-1}\mathbf{B}_e'$, and $\frac{1}{T^2}\sum_t^T \mathbf{X}_t \boldsymbol{\Omega}_{uu}^{-1}\mathbf{X}_t' \xrightarrow{D} \int \mathbf{B}_e \boldsymbol{\Omega}_{uu}^{-1}\mathbf{B}_e'$, a test of the null hypothesis $H_0 : \mathbf{R}\underline{\hat{\beta}}_{\text{dsur}} = \underline{r}$, can be conducted with the Wald statistic

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

which has a limiting $\chi_q^2$ distribution. ‖

*Proof of Proposition* 2. The GLS estimator of $\underline{\beta}$ in (5) can be obtained in two steps. In step 1 we obtain the GLS (or SUR) residual $\underline{\hat{y}}_t$ from the regression of $\underline{y}_t$ on $\mathbf{Z}_{pt}$ and the GLS residual $\hat{\mathbf{X}}_t$ from the regression of $\mathbf{X}_t$ on $\mathbf{Z}_{pt}$. In step 2, we run the GLS (SUR) regression of $\underline{\hat{y}}_t$ on $\hat{\mathbf{X}}_t$. If in step 1 the same set of regressors are used, those GLS regressions are numerically identical to equation-by-equation OLS. ‖

*Proof of Proposition* 3. Follows straightforwardly along the lines of the proof of Proposition 1 and is omitted.

*Proof of Proposition* 4. We begin with $T_*(\underline{\hat{\beta}}_{\text{sysdols}} - \underline{\beta}) = \left(\frac{1}{T_*^2}\sum_{t=p+1}^{T-p}\tilde{\mathbf{X}}_t\tilde{\mathbf{X}}_t'\right)^{-1}\left(\frac{1}{T_*}\sum_{t=p+1}^{T-p}\tilde{\mathbf{X}}_t\underline{u}_t\right)$. By 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 $\mathbf{B}_e$, $T\left(\underline{\hat{\beta}}_{\text{sysdols}} - \underline{\beta}\right) \sim N(0, \mathbf{V}_{\text{sysdols}})$ where $\mathbf{V}_{\text{sysdols}} = \left(\int \mathbf{B}_e\mathbf{B}_e'\right)^{-1}\left(\int \mathbf{B}_e\boldsymbol{\Omega}_{uu}\mathbf{B}_e'\right)\left(\int \mathbf{B}_e\mathbf{B}_e'\right)^{-1}$. The asymptotic chi-square distribution of the Wald statistic follows immediately from the mixed normality of the estimator.

To prove Proposition 5, we make use of the following two lemmas.

**Lemma 1.**

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

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

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

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

$$\text{Var}(Y) = \text{E}[\text{Var}(Y \mid X)] + \text{Var}[\text{E}(Y \mid X)], \tag{A.4}$$

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

**Lemma 2.** *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: $\text{P}(\underline{\lambda}'(\mathbf{A}-\mathbf{B})\underline{\lambda} < 0) > 0$. Then there exists an $\epsilon > 0$ such that $\text{P}(\underline{\lambda}'(\mathbf{A}-\mathbf{B})\underline{\lambda} < -\epsilon) > 0$. 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 $\lim_T \text{P}(\underline{\lambda}'(\mathbf{A}_T - \mathbf{B}_T)\underline{\lambda} < -\delta) = \text{P}(\underline{\lambda}'(\mathbf{A}-\mathbf{B})\underline{\lambda} < -\delta) > 0$, which is a contradiction. ‖

*Proof of Proposition* 5. Let

$$\mathbf{x}_t = \text{diag}(\underline{x}_{1t}, \underline{x}_{2t}) : (2k \times 2), \quad \mathbf{X}_T = \text{diag}(\mathbf{x}_{p+1}, \ldots, \mathbf{x}_{T-p}) : (2T_*k \times 2T_*)$$
$$\mathbf{V}_{1T_*} = \frac{\mathbf{X}_{T_*}'\mathbf{\Omega}^{-1}\mathbf{X}_{T_*}}{T_*^2} \quad \mathbf{V}_{2T_*} = \left(\frac{\mathbf{X}_{T_*}'\mathbf{X}_{T_*}}{T_*^2}\right)\left(\frac{\mathbf{X}_{T_*}'\mathbf{\Omega}\mathbf{X}_{T_*}}{T_*^2}\right)^{-1}\left(\frac{\mathbf{X}_{T_*}'\mathbf{X}_{T_*}}{T_*^2}\right).$$

Then

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

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 $2T_*$ non-negative 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 2, we have $\mathbf{V}_1^{-1} \leq \mathbf{V}_2^{-1}$, and Lemma 1 gives $\text{avar}(\underline{\hat{\beta}}_{\text{dsur}}) \leq \text{avar}(\underline{\hat{\beta}}_{\text{sysdols}})$. ‖

*Comments about SSE*

Here, we show that SSE, the Saikkonen (1993) system estimator, is not efficient for finite samples under regularity conditions of Ogaki and Choi (2001) which we state as:

**Assumption 4.** *(Strict exogeneity with finite leads and lags for a system of dynamic cointegrating regressions)*

(i) *(Strict exogeneity with finite leads and lags). For $u_{it}^{\dagger} = \underline{z}_{it}'\underline{\delta}_i + u_{it}$, which is the projection of the equilibrium error onto $\underline{z}_{pit}' = (\Delta\underline{x}_{1t-p_{1i}}', \ldots, \Delta\underline{x}_{1+q_{1i}}', \ldots, \Delta\underline{x}_{Nt-p_{Ni}}', \ldots, \Delta\underline{x}_{Nt+q_{Ni}}')'$, where $\underline{\delta}_i$ is the vector of unknown projection coefficients, and $\underline{X}_T = [\underline{x}_{11}', \ldots, \underline{x}_{1T}', \ldots, \underline{x}_{N1}', \ldots, \underline{x}_{NT}']'$ with $\underline{x}$ a vector of real numbers,*

$$E(u_{it} \mid \underline{X}_T = \underline{x}) = \underline{0}.$$

(ii) *(Known covariance structure). $E(\underline{u}\underline{u}' \mid \underline{X}_T = \underline{x}) = \mathbf{V}$, where $\underline{u} = (u_{11}, \ldots, u_{1T}, \ldots, u_{N1}, \ldots, u_{NT})'$ and $\mathbf{V}$ is a known non-singular matrix of real numbers.*

Letting $N = 2$, consider the stacked model,

$$\begin{bmatrix}\underline{y}_1\\ \underline{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{W}_1 & \\ & \mathbf{W}_1\end{bmatrix}\begin{bmatrix}\underline{\beta}_1\\ \underline{\delta}_1\\ \underline{\beta}_2\\ \underline{\delta}_2\end{bmatrix} + \begin{bmatrix}\underline{u}_1\\ \underline{u}_2\end{bmatrix} \tag{A.5}$$

where $\mathbf{W}_i = \begin{bmatrix}\underline{x}_{i1}' & \underline{z}_{i1}'\\ \vdots & \vdots\\ \underline{x}_{iT}' & \underline{z}_{iT}'\end{bmatrix}$. Write the stacked model in (A.5) compactly as $\underline{y} = \mathbf{W}\underline{\gamma} + \underline{u}$. We also require:

**Assumption 5.** *For $\underline{X}_T = \underline{x}$, the realized value of* **W** *has full column rank.*

The GLS estimator of the grand coefficient vector $\underline{\gamma}$ is $[\mathbf{W}'\mathbf{V}^{-1}\mathbf{W}]^{-1}\mathbf{W}'\mathbf{V}^{-1}\underline{y}$. By Ogaki and Choi (2001), this GLS estimator is the Best Linear Unbiased Estimator (BLUE) conditional on $\underline{X}_T = \underline{x}$. The DSUR estimator of the cointegration vector parameters is that part of the GLS estimator for $\underline{\beta}$.

To show that SSE is, in this environment, inefficient in finite samples, we need only demonstrate the non-equivalence between SSE and DSUR in the context that we study. We do this for $N = 2$, and $\mathbf{A} = \mathbf{A}' = \mathbf{C} = \text{diag}[a_{11}, a_{22}]$ (a single regressor in each equation) for which we have $\mathbf{B} = \mathbf{I}_2 = [\underline{b}_1, \underline{b}_2]$, $\delta_1 = a_{11}$, $\delta_2 = a_{22}$, $\mathbf{H}_1 = \begin{pmatrix} 1 & 0 \end{pmatrix}'$, $\mathbf{H}_2 = \begin{pmatrix} 0 & 1 \end{pmatrix}'$, $\underline{\hat{a}}_1 = [\hat{a}_{11}, \hat{a}_{12}]'$, $\underline{\hat{a}}_2 = [\hat{a}_{21}, \hat{a}_{22}]'$. For fixed $T$, let $\mathbf{Q}_T = \sum \underline{x}_t \underline{x}_t'$, and notice that $\mathbf{H}'(\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{Q}_T)\mathbf{H} = \mathbf{\Omega}_{uu}^{-1} \odot \mathbf{Q}_T$, where $\mathbf{H} = \text{diag}(\mathbf{H}_1, \mathbf{H}_2)$ and "$\odot$" is the Hamadad product operator.[13] Let $\Omega_{uu}^{ij}$ be the $ij$-th element of $\mathbf{\Omega}_{uu}^{-1}$. In the case where endogeneity is purged by including a fixed number of leads and lags of $\Delta x_{ijt}$, SSE is for fixed $T$

$$\begin{aligned}
\underline{\hat{\beta}}_{\text{SSE},T} &= (\mathbf{H}'(\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{Q}_T)\mathbf{H})^{-1}\mathbf{H}'(\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{Q}_T) vec(\hat{\mathbf{A}}_T') \\
&= (\mathbf{H}'(\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{Q}_T)\mathbf{H})^{-1}\mathbf{H}'(\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{Q}_T)(\mathbf{I} \otimes \hat{\mathbf{Q}}_T^{-1}) vec\left(\sum \underline{\hat{x}}_t \underline{\hat{y}}_t'\right) \\
&= (\mathbf{\Omega}_{uu}^{-1} \odot \mathbf{Q}_T)^{-1}\mathbf{H}'(\mathbf{\Omega}_{uu}^{-1} \otimes \mathbf{Q}_T\, \hat{\mathbf{Q}}_T^{-1}) vec\left(\sum \underline{\hat{x}}_t \underline{\hat{y}}_t'\right) \\
&\neq \underline{\hat{\beta}}_{\text{dsur},T}
\end{aligned}$$

where $\widehat{x}_{it}$ ($\widehat{y}_{jt}$) is the residual from regressing $x_{it}$ ($y_{jt}$) on the leads and lags of $\Delta x_{ijt}$. Clearly the two estimators are not identical for any finite $T$. They are, however, asymptotically equivalent. This can be seen by noting that $T^{-2}(\mathbf{Q}_T - \hat{\mathbf{Q}}_T^{-1}) \to 0$ as $T \to \infty$.

*Acknowledgements*. We thank Ling Hu and Peter Phillips for helpful discussions. For useful comments on earlier versions, we thank seminar participants at the University of California at Santa Barbara, the University of Southern California, the 2003 NBER Summer Institute meeting of the Working Group on Forecasting and Empirical Methods in Macroeconomics and Finance, Mark Watson, two anonymous referees, Hidehiko Ichimura, and Bernard Salanié (the editors). Mark thanks the National Science Foundation for support.

REFERENCES

BAI, J. and NG, S. (2002), "Determining the Number of Factors in Approximate Factor Models", *Econometrica*, **70** (1), 191–221.

BAXTER, M. and CRUCINI, M. (1993), "Explaining Saving–Investment Correlations", *American Economic Review*, **83** (3), 416–436.

BAYOUMI, T. (1997) *Financial Integration and Real Activity* (Ann Arbor: University of Michigan Press).

CANTOR, R. and MARK, N. C. (1988), "The International Transmission of Real Business Cycles", *International Economic Review*, **29** (3), 493–507.

COAKLEY, J. and KULASI, F. (1997), "Cointegration of Long Run Saving and Investment", *Economics Letters*, **54**, 1–6.

COAKLEY, J., KULASI, F. and SMITH, R. (1996), "Current Account Solvency and the Feldstein–Horioka Puzzle", *Economic Journal*, **106**, 620–627.

COAKLEY, J., KULASI, F. and SMITH, R. (1998), "The Feldstein–Horioka Puzzle and Capital Mobility: A Review", *International Journal of Finance and Economics*, **3**, 169–188.

COLE, H. and OBSTFELD, M. (1991), "Commodity Trade and International Risk Sharing: How Much Do Financial Markets Matter?", *Journal of Monetary Economics*, **28** (1), 3–24.

EVANS, M. D. D. and LEWIS, K. K. (1995), "Do Long-Term Swings in the Dollar Affect Estimates of the Risk Premia?", *Review of Financial Studies*, **8** (3), 709–742.

FELDSTEIN, M. and HORIOKA, C. (1980), "Domestic Saving and International Capital Flows", *Economic Journal*, **90** (2), 314–329.

HUSSEIN, K. A. (1998), "International Capital Mobility in OECD Countries: The Feldstein–Horioka 'Puzzle' Revisited", *Economics Letters*, **59**, 237–242.

JANSEN, J. W. (1996), "Estimating Saving–Investment Correlations: Evidence for OECD Countries Based on an Error Correction Model", *Journal of International Money and Finance*, **15** (5), 749–781.

JO, S. and PARK, J. Y. (1997), "On the Comparison of OLS and GLS in Cointegration Regressions", *Journal of Economic Theory and Econometrics*, **3** (2), 1–12.

13. That is, element by element multiplication; $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \odot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{bmatrix}$.

KAO, C. and CHIANG, M.-H. (2000), "On the Estimation and Inference of a Cointegrated Regression in Panel Data", *Advances in Econometrics*, **15**, 179–222.

MADDALA, G. S. and WU, S. (1999), "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test", *Oxford Bulletin of Economics and Statistics*, **61** (suppl. 1), 631–652.

MARK, N. C. and SUL, D. (2003), "Cointegration Vector Estimation by Panel DOLS and Long-Run Money Demand", *Oxford Bulletin of Economics and Statistics*, **65**, 655–680.

MOON, R. H. (1999), "A Note on Fully-Modified Estimation of Seemingly Unrelated Regressions Models with Integrated Regressors", *Economics Letters*, **65**, 25–31.

MOON, R. H. and PERRON, B. (2000), "The Seemingly Unrelated Dynamic Cointegration Regression Model and Testing for Purchasing Power Parity" (Mimeo, University of Southern California).

MOON, R. H. and PERRON, B. (2004), "Testing for a Unit Root in Panels with Dynamic Factors", *Journal of Econometrics*, **122**, 81–126.

OBSTFELD, M. (1986), "Capital Mobility in the World Economy: Theory and Measurement", *Carnegie–Rochester Conference Series on Public Policy*, **24**, 55–104.

OGAKI, M. and CHOI, C.-Y. (2001), "The Gauss–Markov Theorem and Spurious Regressions" (Working Paper No. 02-01, Ohio State University).

PARK, J. Y. (1992), "Canonical Cointegrating Regressions", *Econometrica*, **60** (1), 119–143.

PARK, J. Y. and OGAKI, M. (1991), "Seemingly Unrelated Canonical Cointegrating Regressions" (Working Paper No. 280, Rochester Center for Economic Research).

PARK, J. Y. and PHILLIPS, P. C. B. (1988), "Statistical Inference in Regressions with Integrated Processes: Part 1", *Econometric Theory*, **4**, 468–497.

PEDRONI, P. (2000), "Fully Modified OLS for Heterogeneous Cointegrated Panels", *Advances in Econometrics*, **15**, 93–130.

PHILLIPS, P. C. B. and HANSEN, B. E. (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes", *Review of Economic Studies*, **57**, 99–125.

PHILLIPS, P. C. B. and LORETAN, M. (1991), "Estimating Long-Run Economic Equilibria", *Review of Economic Studies*, **58**, 407–436.

PHILLIPS, P. C. B. and MOON, H. R. (1999), "Linear Regression Limit Theory for Nonstationary Panel Data", *Econometrica*, **67** (5), 1057–1111.

PHILLIPS, P. C. B. and PARK, J. Y. (1988), "Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares in Regressions with Integrated Regressors", *Journal of the American Statistical Association*, **83** (401), 111–115.

PHILLIPS, P. C. B. and SUL, D. (2003), "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence", *The Econometrics Journal*, **6**, 217–260.

SAIKKONEN, P. (1991), "Asymptotically Efficient Estimation of Cointegration Regressions", *Econometric Theory*, **7**, 1–21.

SAIKKONEN, P. (1993), "Estimation of Cointegration Vectors with Linear Restrictions", *Econometric Theory*, **9**, 19–35.

STOCK, J. H. and WATSON, M. W. (1993), "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems", *Econometrica*, **61**, 783–820.