Rochester Center for Economic Research

Seemingly Unrelated Canonical Cointegrating Regressions

Joon Y. Park Masao Ogaki

Working Paper No. 280
June 1991

University of
Rochester

Seemingly Unrelated Canonical Cointegrating Regressions

Joon Y. Park and Masao Ogaki

Rochester Center for Economic Research
Working Paper No. 280

June 1991

Seemingly Unrelated Canonical Cointegrating Regressions

Joon Y. Park
Cornell University

and

Masao Ogaki
University of Rochester

and

The Rochester Center for Economic Research
Working Paper No. 280

Abstract

The paper develops an efficient statistical procedure for the system of seemingly unrelated cointegrating regressions. The method of canonical cointegrating regression for the single cointegrating regression is extended to deal with such a system. The standard procedure for the seemingly unrelated regressions does not necessarily improve upon the single equation least squares, though the system estimation offers an obvious potential for efficiency gain. The potential comes from two different sources: the system-wise information on the presence of unit roots in individual series, and the system error covariance structure. The procedure proposed in the paper fully utilizes both aspects of the system estimation in the seemingly unrelated cointegrating regressions.

First Draft: March 1991

1. Introduction

Various statistical procedures for cointegrated models have rapidly been developed for the past few years. The optimal procedures that are now available to applied researchers include Phillips (1988, 1991), Johansen (1988, 1989), Phillips and Hansen (1990), Park (1991) and Stock and Watson (1991). Though the models that they consider vary, all these different methods yield asymptotically efficient estimates for the cointegrating vectors in general cointegrated models. Also, the problem of nuisance parameter dependency of the tests on the cointegrating vectors, which was addressed in Phillips and Durlauf (1986) and Park and Phillips (1988, 1989), has now been completely solved. The aforementioned procedures all yield asymptotic chi-square tests.

This paper contributes to the above literature by developing an optimal statistical procedure for the system of *seemingly unrelated cointegrating regressions* (SUCR). The SUCR system is just the system of *seemingly unrelated regressions* (SUR) considered in Zellner (1962), with each regression representing a cointegrating relationship. Such a model naturally arises in some important economic applications. For the concrete examples of such models, the reader is referred to Ogaki and Park (1990), Ogaki (1990) and Park and Sung (1989). It appears that there are numerous other econometric applications, for which such modeling deems appropriate and necessary.

None of the existing methods is applicable for inference in the SUCR system. The methods developed by Phillips (1988, 1991), Phillips and Hansen (1990) and Stock and Watson (1991) are for the cointegrated model formulated as a multivariate regression. The typical SUR system is highly overidentified, unlike the standard multivariate regression model which is just identified. As we explain in the paper, the presence of overidentifying restrictions makes the statistical theory of the SUCR system rather different from that of the just identified cointegrated models. The Johansen's (1988, 1989) procedure based on VAR does not allow for a priori identification of individual cointegrating relationships, and can be used only for the estimation of just identified models. It is, in particular, not applicable for the SUCR system.

The method that we propose in the paper is based on the *canonical cointegrating regression* (CCR), which was recently developed by Park (1991). The CCR procedure was originally proposed for the estimation of and testing in a single equation. We extend the procedure here to a system of cointegrating regressions that are seemingly unrelated. Naturally, we call the CCR procedure applied to the SUCR system the *seemingly unrelated canonical cointegrating regression* (SUCCR) procedure. As in the CCR procedure, the construction of a SUCCR requires the transformation of a system using stationary components of the model, so that the usual least squares method yields an efficient parameter estimate and asymptotic chi-square tests on the coefficients, if based on the transformed model.

Although the methods of inference in cointegrated models by Phillips (1988, 1991), Johansen (1988, 1989), Phillips and Hansen (1990) and Stock and Watson (1991) were developed explicitly in multivariate framework so as to deal with more than one longrun relationships, they can be used only for models with no overidentifying restrictions in cointegrating coefficients. As noted in Park (1991), the system estimation has no comparative advantage in such models, over the separate single equation estimation of each cointegrating relationship. In particular, the single equation CCR procedure is asymptotically equivalent to all the other aforementioned methods that are based on models including multiple cointegrating relationships.

The SUCCR procedure improves in two ways upon the single equation CCR or any of the existing efficient methods applied to a given SUCR system equation by equation. First, the SUCCR transformation utilizes the presence of the unit roots in the entire system, contrary to the single equation efficient methods which rely only on the unit roots in an equation. The additional information on the unit roots exploitable through the system estimation is roughly proportional to the degree of overidentification of a given equation. Clearly, the number of independent unit roots in each equation is identical to that of the entire system for just identified models (The unit root in the regressand is implied by the unit roots in the regressors and the presence of cointegration between the regressand and the regressors). For such a model, no efficiency gain is expected from the system information on the unit roots. In

sharp contrast, the system estimation has the greatest potential to improve efficiency in the standard SUCR system with nonoverlapping regressors in each equation.

Second, the SUCCR estimator further improves efficiency through the application of the system GLS. Just as in the standard regression model, the GLS procedure utilizing information on the covariance structure of the system errors results in improved efficiency in the SUCR system. This may well seem to contradict Phillips and Park (1988), which shows the equivalence of OLS and GLS in cointegrating regression. It is made clear in this paper, however, that their equivalence result for single cointegrating regressions does not extend to a system of cointegrating equations. In the SUCR system, the system GLS has an obvious potential to improve upon OLS. The SUCCR method utilizes the potential, and is unambiguously more efficient than the OLS procedure applied to the same transformed system.

The Monte Carlo simulation is performed to compare the finite sample performance of the system estimation with the single equation methods. The efficiency gain by the SUCCR procedure is well evidenced in the simulation. Even for the samples of moderate size, it seems apparant that we may benefit much from the use of the system SUCCR procedure. The SUCCR method substantially improves upon the single equation methods, in virtually all the cases that we investigate, in terms of both bias and mean square error. Both the OLS and GLS estimators in the SUCCR system are considered in the simulation. The former only utilizes the system-wise information on the unit roots, while the latter also uses the covariance structure of the system errors. As one might well expect, the GLS estimator using full system information performs largely better in finite samples than the OLS estimator relying only on a partial information.

The rest of the paper is organized as follows: Section 2 explains the models and assumptions. The statistical theory of the standard SUCR system is presented in Section 3. The limiting distributions of the least squares estimators in the SUCR system are analyzed, with focus on the presence of bias and non-normality, and on the efficiency of GLS relative to OLS. The simple prototype model is considered in detail to separate various issues. Section 4 introduces the SUCCR transformation

and explains its motivation. The statistical theory of the SUCCR estimator is also developed there. The feasible SUCCR procedure, which requires pre-estimation of the various nuisance parameters, is given in Section 5. Section 6 includes the simulation results on the finite sample performance of the SUCCR estimator and others. Section 7 concludes the paper. Mathematical proofs are in Appendix.

2. The Models and Assumptions

We consider a system of regressions given by

$$\begin{aligned} y_{1t} &= x_{1t}'\beta_1 + u_{1t} \\ y_{2t} &= x_{2t}'\beta_2 + u_{2t} \\ &\vdots \quad\quad \vdots \\ y_{mt} &= x_{mt}'\beta_m + u_{mt} \end{aligned} \tag{1}$$

To write the system with n-observations in matrix form, let $y_i = (y_{i1}, \ldots, y_{in})'$, $X_i = (x_{i1}, \ldots, x_{in})'$ and $u_i = (u_{i1}, \ldots, u_{in})'$ for $i = 1, \ldots, m$. Moreover, we subsequently define $y = (y_1', \ldots, y_m')'$, $\beta = (\beta_1', \ldots, \beta_m')'$, $u = (u_1', \ldots, u_m')'$, and finally, X to be a block diagonal matrix with X_i in the i-th diagonal for $i = 1, \ldots, m$. The system (1) now becomes

$$y = X\beta + u$$

in matrix form.

We primarily assume in the paper that $\{x_{it}\}$, $i = 1, \ldots, m$, are integrated processes (of order one), which may include, explicitly or implicitly, deterministic trends such as time polynomials and dummy variables. More precisely, we specify $\{x_{it}\}$ in the following three different ways:

$$\begin{aligned} \mathrm{M}(a): \quad x_{it} &= x_{it}^0 \\ \mathrm{M}(b): \quad x_{it} &= \pi_i p_{it} + x_{it}^0 \\ \mathrm{M}(c): \quad x_{it} &= (p_{it}', q_{it}')', \qquad q_{it} = \pi_i p_{it} + x_{it}^0 \end{aligned}$$

where $\{p_{it}\}$ is a general deterministic trend and $\{x_{it}^0\}$ is a purely stochastic integrated process. The residuals $\{u_{it}\}$ are assumed to be stationary.

Under these specifications, our model (1) roughly describes a system of cointegrating regressions. The precise meaning of the i-th relationship in (1) is, however, different for each of the specifications of $\{x_{it}\}$ in M(a) - M(c). With M(a), the i-th regression in (1) represents cointegration in the sense of Engle and Granger (1987). The regressor $\{x_{it}\}$ does not have any deterministic component. Clearly, $\{y_{it}\}$ should also be purely stochastic with this specification of $\{x_{it}\}$ for the i-th relationship in (1) to hold.

Under the specification in M(b), both $\{y_{it}\}$ and $\{x_{it}\}$ contain the deterministic trend $\{p_{it}\}$. The i-th relationship in (1) in this case is stronger than is implied by M(a). To see this more clearly, we let $\{y_{it}\}$ be generated as $y_{it} = \tau_i' p_{it} + y_{it}^0$. It follows then directly from (1) that

$$y_{it}^0 = x_{it}^{0\prime}\beta_i + u_{it} \tag{2}$$

and

$$\tau_i - \pi_i'\beta_i = 0 \tag{3}$$

The i-th equation thus describes co-trending of the deterministic components, as well as the usual cointegration in the stochastic components. Such relationship is called *deterministic cointegration* in the paper.

When $\{x_{it}\}$ is given as in M(c), $\{y_{it}\}$ may or may not have a deterministic trend. The inclusion of the deterministic trend $\{p_{it}\}$ in the regression effectively detrends both series in any case. The i-th relationship in (1) therefore describes cointegration between only the stochastic components of $\{y_{it}\}$ and $\{x_{it}\}$, which we may call *stochastic cointegration.* In the separate formulation of the stochastic and the deterministic relationships in (2) and (3), only the former is required to hold. The relationship (3) may also hold, but in this case the regression formulated as in M(b) is more appropriate, and yields more efficient estimate of β_i.

We consider in this paper a system of cointegrating regressions specified by any of M(a) - M(c). Any mixture of these three types of cointegrating relationships

is allowed. The class of models is broad enough to cover a system with equations individually representing any cointegrating relationship that has ever been looked at in the literature. The method proposed in the paper is in fact flexible enough to deal with a mixture of cointegrating regressions and the usual stationary regressions. We may indeed allow for a system involving equations with

$$\mathrm{M}(d): \quad x_{it} = p_{it}$$

in addition to those with M(a) - M(c). Under the specification of $\{x_{it}\}$ as M(d), $\{y_{it}\}$ becomes trend stationary and the i-th regression becomes a standard stationary regression.

The longrun relationships in (1) with any of the specifications M(a) - M(c) are testable, through testing for cointegration. Here and in many other contexts where the presence, not the absence, of cointegration is postulated, it seems preferable to test for the null hypothesis of cointegration. The test of cointegration based on variable addition is available in Park (1990a) and Park, Ouliaris and Choi (1988). All the other existing tests take non-cointegration as the null hypothesis. If such tests are used, the null hypothesis must be rejected to validate our specification (1). This would make it difficult to interprete the size of the tests, intended here to check the adequacy of model specification. The residual based tests of non-cointegration proposed in Engle and Granger (1987) and analyzed later in Phillips and Ouliaris (1990), and the tests by Stock and Watson (1987) and Johansen (1988, 1989) on the number of cointegration can be more useful in some other contexts, as we will explain later.

We now define

$$w_t = (u_t', \Delta x_t^{0\prime})' \tag{4}$$

where $u_t = (u_{1t}, \ldots, u_{mt})'$ and $\Delta x_t^0 = (\Delta x_{1t}^{0\prime}, \ldots, \Delta x_{mt}^{0\prime})'$. The vector of stationary processes driving the system is thus denoted by $\{w_t\}$. We assume throughout the paper that $\{w_t\}$ satisfies an invariance priciple. That is,

$$B_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} w_t \xrightarrow{\mathcal{D}} B(r) \tag{5}$$

where B is a vector Brownian motion. In (5) and elsewhere in the paper, $[z]$ denotes the largest integer which does not exceed z.

The invariance principle (5) holds under very general conditions, and is valid for $\{w_t\}$ driven by a large class of models including virtually all that are practically important. Likewise, the deterministic trend $\{p_{it}\}$ in M(a) - M(d) is only required to satisfy some fairly mild regularity conditions. It will be sufficient to assume for each component $\{p_{ijt}\}$ of $\{p_{it}\}$ that $p_{ijt}/n^{\delta_{ij}}$ for some δ_{ij} has a properly defined limit in $L^2[0,1]$ of square integrable functions. More precisely, we let

$$f_{in}(r) \;=\; \Delta_{in}^{-1} p_{i[nr]} \;\xrightarrow{\mathcal{L}^2}\; f_i(r) \tag{6}$$

where Δ_{in} is a diagonal matrix with the j-th diagonal element $n^{\delta_{ij}}$. Popular stationary ARMA models with deterministic trends consisting of polynomials in time and dummy variables, of course, meet all the required conditions. For the explicit conditions for $\{w_t\}$ and $\{p_{it}\}$, the reader is referred to Park (1991) and the references cited there.

The covariance matrix Ω of the limit Brownian motion B in (5) is given by

$$\Omega = \lim_{n\to\infty} \frac{1}{n} E \left(\sum_{t=1}^{n} w_t \right) \left(\sum_{t=1}^{n} w_t \right)' = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \tag{7}$$

where the partition is made conformably with that of $\{w_t\}$ in (4). We call Ω in (7) the *longrun* variance of $\{w_t\}$ in this paper. The usual variance of $\{w_t\}$ is denoted by Σ, i.e.,

$$\Sigma = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E(w_t w_t') = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{8}$$

with the partition made similarly as that of the longrun variance Ω. The usual and the longrun variances Σ_{11} and Ω_{11} of the system errors will also be denoted by Σ_0 and Ω_0, respectively. When $\{w_t\}$ is a martingale difference sequence, Ω becomes identical to Σ.

For the equation specified by M(b), it is assumed with no loss in generality that each component of the deterministic trend $\{p_{it}\}$ is of order greater than $\sqrt{n}$. For instance, $p_{it} = t$ is such a deterministic trend since it is of order n, while $p_{it} = 1$ with

order 1 is not. All the deterministic trends in $\{x_{it}\}$ of order smaller than $\sqrt{n}$ are absorbed in $\{x_{it}^0\}$ and become unimportant in asymptotics. Moreover, we assume, by ignoring lower order terms in $\{p_{it}\}$ if necessary, that π_i is of full column rank. Under these innocuous assumptions, we define an orthogonal matrix $H_i = (H_{1i}, H_{2i})$ such that

$$H_{1i}'\pi_i = I \quad \text{and} \quad H_{2i}'\pi_i = 0 \tag{9}$$

with the convention $H = H_{1i}$ when π_i is of full rank. In our subsequent results, all terms involving H_{2i} and the expressions corresponding to them vanish in this case.

We assume throughout the paper that

$$\Omega_{22} > 0 \tag{10}$$

which implies that there are neither redundant nor cointegrated variables in $\{x_t\}$. Violation of (10) in the system (1) may arise in two different contexts. First, the condition (10) is violated, when the individual components of the regressor $\{x_{it}\}$ are cointegrated for any $i = 1, \ldots, m$. This causes a serious problem for equations specified as M(a) and M(c), since we then have more than one cointegrating vectors in $\{z_{it}^0\}$, where $z_{it}^0 = (y_{it}^0, x_{it}^{0\prime})'$, and the coefficients for the stochastic regressors are not uniquely determined. We exclude this possibility. The residual based tests by Engle and Granger (1987) and Phillips and Ouliaris (1990) can be used directly to test for non-cointegration of the stochastic components in $\{x_t\}$. For equations specified by M(b), this problem of the lack of identifiability does not arise, as long as the leading deterministic terms (of order higher than $\sqrt{n}$) are distinct.

Second, we may also encounter a model in which $\{x_{it}\}$ and $\{x_{jt}\}$ for $i \neq j$ have some common and/or crossly cointegrated variables. Such a system is allowed in our subsequent analysis. Clearly, the condition (10) can always be made fulfilled by considering a proper subset of $\{x_t\}$, which does not include any redundant or cointegrated variables. Any redundant variables can simply be thrown away. Also, with the help of the tests for the number of cointegrations by Stock and Watson (1988) and Johansen (1988, 1989) and subsequently using tests for the presence and absence of cointegration, we may exclude as many variables as the number of coin-

tegrations to select the appropriate subset. Our proposed method is valid with the corresponding redefinition of $\{\Delta x_t^0\}$ in (4), and Ω in (7). The system with crossly cointegrated regressors is incomplete, however, since the cointegrations between regressors provide additional independent longrun relationships. Therefore, estimating such models generally yields sub-optimal coefficient estimates.

Finally, it will be convenient to define B_1 and B_2 to be the limit Brownian motions corresponding to $\{u_t\}$ and $\{\Delta x_t^0\}$, respectively, and B_{2i} to be the Brownian motion representing $\{\Delta x_{it}^0\}$. Then we let

$$U = B_1 \tag{11}$$

and V be a block diagonal matrix with the i-th entry

$$V_i = B_{2i}, \left(\begin{array}{c} f_i \\ H_{2i}' B_{2i} \end{array} \right), \left(\begin{array}{c} f_i \\ B_{2i} \end{array} \right), f_i \tag{12}$$

for the i-th equation given by M(a) - M(d), respectively.

3. Statistical Theory of SUR

The SUR estimator of β in (1) is given by

$$\hat{\beta}_{\text{SUR}} = (X'(\hat{\Sigma}_0^{-1} \otimes I)X)^{-1} X'(\hat{\Sigma}_0^{-1} \otimes I)y \tag{13}$$

where $\hat{\Sigma}_0 = \sum_{t=1}^n \hat{u}_t \hat{u}_t'/n$, i.e., the usual variance estimate of $\{u_t\}$ based on the fitted residual $\{\hat{u}_t\}$. The system GLS estimator $\hat{\beta}_{\text{SUR}}$ is compared with the OLS estimator $\hat{\beta}_{\text{OLS}}$ defined as

$$\hat{\beta}_{\text{OLS}} = (X'X)^{-1} X'y \tag{14}$$

As is well known, the system OLS is identical to the single equation OLS in (1).

To represent the limiting distributions of $\hat{\beta}_{\text{SUR}}$ and $\hat{\beta}_{\text{OLS}}$, defined in (13) and (14), it will be convenient to introduce some additional notation. Let D_n be a block diagonal matrix with the i-th diagonal entry

$$D_{in} = \sqrt{n} I, \left(\begin{array}{cc} \Delta_{in} & 0 \\ 0 & \sqrt{n} I \end{array} \right), \left(\begin{array}{cc} \Delta_{in} & 0 \\ 0 & \sqrt{n} I \end{array} \right), \Delta_{in} \tag{15}$$

for the i-th equation given respectively by M(a) - M(d), where Δ_{in} is defined in (6). Similarly, T is a block diagonal matrix with

$$T_i = I, \ H_i', \ \begin{pmatrix} I & -\pi_i' \\ 0 & I \end{pmatrix}, \ I \tag{16}$$

for each of the equations specified as M(a) - M(d), where H_i is given in (9). The dimensions of the matrices and submatrices in (15) and (16) are conformable to those in (12) for each case.

Moreover, we let

$$\lambda_{ij} = \lambda_{ij}^0, \ \begin{pmatrix} 0 \\ H_{2i}'\lambda_{ij}^0 \end{pmatrix}, \ \begin{pmatrix} 0 \\ \lambda_{ij}^0 \end{pmatrix}, \ 0 \tag{17}$$

for equations given by M(a) - M(d), where

$$\lambda_{ij}^0 = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E(x_{it}^0 u_{jt})$$

Theorem 1 *Let U and V be defined as in* (11) *and* (12), *and let D_n and T be given as above in* (15) *and* (16). *Then*

$$\sqrt{n} D_n T(\hat{\beta}_{\text{SUR}} - \beta) \xrightarrow{\mathcal{D}} \left(\int_0^1 V\Sigma_0^{-1}V'\right)^{-1} \left(\int_0^1 V\Sigma_0^{-1}\, dU + \nu\right)$$
$$\sqrt{n} D_n T(\hat{\beta}_{\text{OLS}} - \beta) \xrightarrow{\mathcal{D}} \left(\int_0^1 VV'\right)^{-1} \left(\int_0^1 V\, dU + \mu\right)$$

The $\nu = (\nu_1', \ldots, \nu_m')'$ and $\mu = (\mu_1', \ldots, \mu_m')'$ are given by $\nu_i = \sum_{j=1}^m \sigma^{ij}\lambda_{ij}$ and $\mu_i = \lambda_{ii}$, where $\Sigma_0^{-1} = (\sigma^{ij})$ and λ_{ij} is defined in (17).

Theorem 1 indicates that the limiting distributions of $\hat{\beta}_{\text{SUR}}$ and $\hat{\beta}_{\text{OLS}}$ are, in general, non-Gaussian. Both estimators are not only inefficient, but also asymptotically biased, in a general SUCR system. The asymptotic bias is generated from two sources: the correlation between the limit processes U and V, and the bias terms ν and μ respectively for the GLS and OLS estimators. Inefficiency results in from not utilizing the presence of the unit roots in individual series and the covariance structure of the system errors, as we will see later in this paper. Similarly as the results in Phillips and Durlauf (1986) and Park and Phillips (1988) for the multivariate cointegrating

regression, the limiting distribution of the usual chi-square tests in a general SUCR system involves nuisance parameters. This invalidates the use of the standard tests.

Moreover, the nonstandard nature of the limiting distributions in Theorem 1 makes it impossible to directly compare the OLS and the GLS procedures in a general SUCR system. Most of all, the usual superiority of GLS is not warranted. There is no reason to believe that the standard system GLS method would improve efficiency, given the non-normality of the limiting distributions of the least squares estimators and the presence of bias terms. Evidently, the GLS estimator in a general SUCR system is not necessarily better than the OLS estimator. There are, in fact, several reasons to believe that GLS can even be worse than OLS, as we will explain later.

To properly understand various issues involved in the estimation of the SUCR system, it seems very useful to consider the *prototype* model in which

PT(a): $\{u_t\}$ is uncorrelated with $\{\Delta x_t^0\}$, *serially* as well as contemporaneously, and

PT(b): $\{u_t\}$ is a martingale difference sequence.

We assume in PT(a) that that the regression errors in the prototype model are uncorrelated, contemporaneously and serially, with the innovations of the stochastic componenets of the regressors. This is stronger than the usual exogeneity condition. We require here strict exogeneity so that all the future, as well as the present and past, innovations of the regressors are uncorrelated with the regression errors. The condition PT(b) ensures that the longrun variance of the errors is Σ_0, and the use of $\hat{\Sigma}_0$ as the weighting matrix for the system GLS is appropriate. This assumption is commonly made in the standard SUR.

Lemma 2 *Consider the prototype model satisfying* PT(a) *and* PT(b). *We have*

$$\sqrt{n}D_nT(\hat{\beta}_{\text{SUR}} - \beta) \xrightarrow{\mathcal{D}} \int_V N(0, P(V, \Sigma_0))\, dV$$

$$\sqrt{n}D_nT(\hat{\beta}_{\text{OLS}} - \beta) \xrightarrow{\mathcal{D}} \int_V N(0, Q(V, \Sigma_0))\, dV$$

where P and Q are random matrices given by

$$P(V,\Sigma_0) = \left(\int_0^1 V\Sigma_0^{-1}V'\right)^{-1}$$
$$Q(V,\Sigma_0) = \left(\int_0^1 VV'\right)^{-1}\int_0^1 V\Sigma_0 V'\left(\int_0^1 VV'\right)^{-1}$$

and other notations are defined in Theorem 1.

The standard SUR estimator for the prototype SUCR is asymptotically efficient and unbiased. When the conditions PT(a) and PT(b) are strengthened so that the errors $\{u_t\}$ are iid with normal distribution and completely independent of $\{\Delta x_t^0\}$, the standard SUR estimator is indeed the exact ML estimator. The generalizations in PT(a) and PT(b) do not affect the asymptotic results. With the strict exogeneity in PT(a), any information on $\{x_t\}$, including the presence of the unit roots, is irrelevant to the efficient estimation of β. This is why the standard SUR is fully efficient even without utilizing the information on the unit roots.

The statistical theory for the prototype SUCR model is largely identical to that of the standard SUR system, mostly due to the mixed normality. It is in fact easy to show that the standard tests, such as Wald, LR and LM, are valid, and have the usual asymptotic chi-square distribution. The tests on the prototype SUCR coefficients β can therefore be done, exactly as in the standard SUR. The reader is referred to Park (1991) for a more detailed discussion on this subject.

Moreover, the mixed normality allows us to directly compare the asymptotic variances of the GLS and OLS procedures. For the protype SUCR model, the system GLS method is, in particular, unambiguously more efficient than the OLS method.

Corollary 3 *Consider the prototype model satisfying* PT(a) *and* PT(b). *We have*

$$\text{avar}\left(\hat{\beta}_{\text{SUR}}\right) \leq \text{avar}\left(\hat{\beta}_{\text{OLS}}\right)$$

in positive definite sense.

The relative efficiency of $\hat{\beta}_{\text{SUR}}$ over $\hat{\beta}_{\text{OLS}}$ is therefore well established for the prototype SUCR model. In Corollary 3 and elsewhere in the paper, 'avar' denotes the asymptotic variance.

It is now evident, at least for the prototype SUCR, that the system GLS does improve upon the OLS procedure. This is in sharp contrast with Phillips and Park (1988), which shows the equivalence of the GLS and OLS procedures for the single cointegrating regression. Considering a cointegrating regression for which the errors are autoregressive processes generated independently from the regressors, they made it clear that the GLS method has no potenetial to improve efficiency in the single equation context. We show here that their finding is not extended to the system estimation.

The relative efficiency of the standard SUR estimator, however, does not apply to a more general SUCR system. For a general system for which the errors are autocorrelated and/or correlated with the innovations of the stochastic regressors, there are reasons to believe that the system GLS method can even have adverse effects. First, the GLS transform can make the errors have a more complicated covariance structure, by improperly weighting the errors using the usual covariance matrix estimate. The correct weighting matrix for a general SUCR is the longrun variance Ω_0, not the usual variance Σ_0, of the errors. Second, the GLS procedure can magnify the asymptotic bias. The bias term (μ for the OLS estimator) may well be inflated (to ν for the GLS estimator) through the GLS transform. Corollary 3 only shows the *potential* of efficiency gain in the SUR, and should not be interpreted generally as the superiority of the standard SUR procedure for SUCR models.

One may easily see from the proofs of Lemma 2 and Corollary 3 that the mixed normality and its consequence on the relative efficiency of the system GLS procedure in the SUCR system (1) rely on the following three condition:

C(a): $\nu = \mu = 0$,
C(b): U is independent of V, and
C(c): U has covariance matrix Σ_0

in the notation used in Theorem 1. The conditions C(a) and C(b) together reduce the limiting distributions of the least squares estimators in Theorem 1 to mixed normal, and C(c) ensures that the SUR estimator more efficient than the OLS estimator. For

the prototype model, C(a) and C(b) follow from PT(a), and C(c) from PT(b).

Of the three conditions, C(c) is much less critical than the other two. Violation of C(c) can easily be dealt with a slight modification of the standard SUR estimator. It is, in fact, not difficult to see that the system GLS estimator defined as

$$\hat{\beta}_{\text{MSUR}} = (X'(\hat{\Omega}_0^{-1} \otimes I)X)^{-1}X'(\hat{\Omega}_0^{-1} \otimes I)y \tag{18}$$

i.e., the SUR estimator with $\hat{\Sigma}_0$ replaced by a consistent estimator of Ω_0, is more efficient than the OLS estimator under only C(a) and C(b). All our previous results for the prototype model hold without the assumption PT(b) for the modified SUR estimator in (18).

The modified SUR estimator is asymptotically normal and relatively more efficient than the OLS estimator, when the model is given by M(d) for all the equations. The regressions in the system would then be the stationary regressions on deterministic trends. Such a system, of course, trivially satisfies the conditions C(a) and C(b), as one may easily see in Theorem 1. Since the limiting matrix V becomes deterministic in this case, the mixed normality is reduced to normality. Rather obviously,

$$\sqrt{n}D_nT(\hat{\beta}_{\text{MSUR}} - \beta) \xrightarrow{\mathcal{D}} N(0, P(V, \Omega_0)) \tag{19}$$

for the system of regressions on deterministic trends.

The system of deterministic cointegrations with specification M(b) may essentially turn into such regressions, at least asymptotically. This happens when π_i in M(b), $i = 1, \ldots, m$, is of full row rank for each equation. In this case, $\{x_{it}\}$ is driven effectively by its deterministic component $\pi_i p_{it}$, which dominates the stochastic component $\{x_{it}^0\}$. As a result, $\{x_{it}\}$ generated by M(b) asymptotically behaves exactly as the deterministic regressor given by M(d). The i-th regression may therefore be treated as a stationary regression. The asymptotic result (19) follows easily from Theorem 1 with our convention on the matrix H in (9). The asymptotic normality and the relative efficiency of the modified SUR estimator, therefore, applies without any assumption on the stochastic component $\{x_t^0\}$ such as PT(a). This generalizes earlier results by Park and Phillips (1988) and West (1988) for multivariate and bivariate regression models, respectively.

For the system with equations specified by M(d), the modified SUR estimator is asymptotically efficient. It is, however, generally *not* fully efficient in the system with equations specified as M(b) with the full row rank condition for π_i, $i = 1, \ldots, m$. Unless the stochastic components of the regressors given by M(b) are asymptotically independent of the system errors as in PT(a), their unit roots may always be used to strictly increase the estimator efficiency in a SUCR. This will be shown in the next section. Here we simply point out that the presence of the unit roots is not exploited in the standard or modified SUR procedure.

The normality in (19) does not apply for a system consisting of equations given by M(b), when π_i has deficiency in row rank for some of the regressions. For such regressions, $\{x_{it}\}$ is not entriely dominated by its deterministic trend, and there exist some linear combinations of $\{x_{it}\}$ that are purely stochastic. The row rank of π_i is unknown, but we must have at least as many distinct determinstic trend terms as the number of regressors, to avoid deficiency in its row rank. For the result (19) to hold, it is required that π_i must be of full row rank for all the equations, even if we are concerned with only a single or a subsystem of equations. Otherwise, the modified SUR has nonnormal limiting distribution, and is not necessarily more efficient than the OLS estimator. The full rank condition is necessary only for the equations of interest, for the OLS estimator to have asymptotic normal distribution.

The equation specified by M(c) does not behave like M(b) or M(d) in any case, though the deterministic trends are included in individual series. This is because we effectively detrend the deterministic components by maintaining the trend in the regression. The statistical theory for a system with equations given by M(c) is similar to that for the purely stochastic model driven by M(a). The exogeneity assumption like PT(a) is therefore crucial for our results on the mixed normality and the relative efficiency of the SUR estimator.

As we have clearly seen, the system estimation in the SUCR system offers a potential to improve upon the single equation estimation. The potential is, however, fully utilized by the standard or the modified SUR procedure only for the models with strictly exogenous regressors or models behaving as such asymptotically. Un-

fortunately, the strict exogeneity condition is unrealistic, and rarely expected to hold in practical applications. The statistical theory for the usual SUR procedure under the assumption of strict exogeneity may therefore seem to have little relevancy in practice. It is, however, very important, from theoretical point of view. It will in fact be shown in the next section that we may always transform a given SUCR system so that it behaves *asymptotically* as the system with strictly exogenous regressors and the theory for the prototype model applies. The method that we will develop in the next section most efficiently uses the system information for a general SUCR model, just as the standard SUR procedure does for the prototype model.

4. SUCCR

In this section, we will extend to the SUCR system the CCR methodology by Park (1991), which was developed originally for single cointegrating regressions. The CCR is a transformed regression formulated in such a way that the usual least squares procedure yields both efficient estimates and chi-square tests on the coefficients. The method utilizes the fact that cointegrating regressions are not unique, and any models which differ only by stationary deviations represent the same cointegrating relationship. For a given cointegrating regression, the CCR method requires the transformation of the data using the stationary components of the model. The presence of the unit roots in individual series is utilized in the transformation.

Consider the transformations of $\{y_{it}\}$ and $\{x_{it}\}$ given for each i by

$$\begin{aligned} y_{it}^{*} &= y_{it} - \omega_{12}^{i}\Omega_{22}^{-1}\Delta x_{t}^{0} - \beta_{i}'\Lambda_{i}\Sigma^{-1}w_{t} \\ x_{it}^{*} &= x_{it} - \Lambda_{i}\Sigma^{-1}w_{t} \end{aligned} \tag{20}$$

where ω_{12}^{i} is the i-th row of Ω_{12},

$$\Lambda_{i} = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}E(x_{it}w_{t}') \tag{21}$$

and other notations are defined earlier in (4), (7) and (8).

Notice that the cointegrating relationships (1) continue to hold for the transformed variables $\{y_{it}^*\}$ and $\{x_{it}^*\}$, $i = 1, \ldots, m$, since the transformation in (20) involves only stationary terms. We now have the transformed model, which we call the SUCCR,

$$\begin{aligned} y_{1t}^* &= x_{1t}^{*\prime}\beta_1 + u_{1t}^* \\ y_{2t}^* &= x_{2t}^{*\prime}\beta_2 + u_{2t}^* \\ &\vdots \quad\quad \vdots \\ y_{mt}^* &= x_{mt}^{*\prime}\beta_m + u_{mt}^* \end{aligned} \tag{22}$$

where $u_{it}^* = u_{it} - \omega_{12}^i \Omega_{22}^{-1} \triangle x_t^0$.

Define $x_t^* = (x_{1t}^{*\prime}, \ldots, x_{mt}^{*\prime})'$ and $u_t^* = (u_{1t}^*, \ldots, u_{mt}^*)'$, similarly as $\{x_t\}$ and $\{u_t\}$. It follows that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E(x_t^* u_t^{*\prime}) = 0 \tag{23}$$

i.e., the orthogonality of the regressors and the system errors holds in the SUCCR system. Moreover, we have

$$u_t^* = u_t - \Omega_{12}\Omega_{22}^{-1} \triangle x_t^0 \tag{24}$$

and the SUCCR errors are asymptotically independent of the regressors.

Let y_*, X_* and u_* be defined respectively from $\{y_{it}^*\}$, $\{x_{it}^*\}$ and $\{u_{it}^*\}$, for $i = 1, \ldots, m$ and $t = 1, \ldots, n$, in the same way as y, X and u for the system (1). We then write the SUCCR (22) in matrix form as

$$y_* = X_*\beta + u_*$$

The SUCCR estimator that we propose to use is just the modified system GLS estimator in (22), using the longrun variance

$$\Omega_* = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21} \tag{25}$$

of the SUCCR errors $\{u_t^*\}$. The SUCCR estimator is given explicitly as

$$\hat{\beta}_{\text{SUCCR}} = (X_*'(\Omega_*^{-1} \otimes I)X_*)^{-1} X_*'(\Omega_*^{-1} \otimes I)^{-1} y_* \tag{26}$$

We also consider the OLS procedure applied to the SUCCR system, which we call the *system* CCR (SCCR) in the paper. The SCCR estimator $\hat{\beta}_{\text{SCCR}}$ is

$$\hat{\beta}_{\text{SCCR}} = (X_*'X_*)^{-1}X_*'y_*$$

The SCCR estimator is, of course, identical to the single equation OLS estimator in the SUCCR system, which is obtained by applying OLS to (22) equation by equation. The SCCR procedure, however, should not be confused with the single equation CCR method in Park (1991). The transformation for the single equation CCR is based on the presence of unit roots only in the series included in each equation. On the contrary, the SUCCR transformation in (20) utilizes the unit roots in the entire system.

The limiting distributions of $\hat{\beta}_{\text{SUCCR}}$ and $\hat{\beta}_{\text{SCCR}}$ can easily be derived from Theorem 1 and the subsequent results on the prototype model.

Theorem 4 *We have*

$$\sqrt{n}D_nT(\hat{\beta}_{\text{SUCCR}} - \beta) \xrightarrow{\mathcal{D}} \int_V N\left(0, P(V, \Omega_*)\right)\, dV$$
$$\sqrt{n}D_nT(\hat{\beta}_{\text{SCCR}} - \beta) \xrightarrow{\mathcal{D}} \int_V N\left(0, Q(V, \Omega_*)\right)\, dV$$

where P and Q are the random matrices defined in Lemma 2 *and other notations are given in* Theorem 1.

The limiting distributions of the least squares estimators in the SUCCR system (22) are mixed normal for a general, as well as for the prototype, SUCR model. This is well expected from the properties (23) and (24) of the SUCCR system. It is indeed not difficult to see that the orthogonality condition (23) ensures the bias term similarly defined as ν or μ in Theorem 1 to vanish. Moreover, the limit process of the SUCCR errors $\{u_t^*\}$ is given from (24) by

$$U_* = B_1 - \Omega_{12}\Omega_{22}^{-1}B_2$$

which is independent of V for models with equations given by any mixture of M(a) - M(d). The conditions corresponding to C(a) and C(b) in Section 3 hold in the SUCCR system (22) for a general SUCR model.

The SUCCR estimator is asymptotically equivalent to the exact ML estimation in a Gaussian parametric ECM model with the cointegrating relationships in (1). The exact ML estimator for the longrun parameters in such an ECM is derived in Park (1990b), to which the interested reader is referred for a detailed discussion on the ML interpretation of the CCR procedure. As we mentioned in the introduction, the Johansen's (1988, 1989) method is not applicable for an ECM including overidentified longrun relationships as those in our SUCR (1).

The regressions in the prototype models are already formulated in canonical form, and the construction of the SUCCR (22) is unnecessary. The transformations (20) are, however, innocuous. The least squares procedures in the SUCCR system are asymptotically equivalent to those in the original model. This is because $\Omega_* = \Sigma_0$ $(\equiv \Sigma_{11})$ for the prototype model. For the system of deterministic cointegrating regressions given by M(b) with the full row rank condition for π_i $i = 1, \ldots, m$, in contrast, the SUCCR transformation changes the system errors. In such systems, the SUCCR estimator has the limiting distribution

$$\sqrt{n} D_n T(\hat{\beta}_{\text{SUCCR}} - \beta) \xrightarrow{\mathcal{D}} N(0, P(V, \Omega_*))$$

since V becomes deterministic. The SUCCR estimator now can easily be seen to be more efficient than the modified SUR estimator in the original model, the limiting distribution of which is given in (19). Notice that $\Omega_* \leq \Omega_0$ $(\equiv \Omega_{11})$.

As one may well expect from our result in Section 3 on the comparison of GLS and OLS for the prototype model, the SUCCR estimator improves upon the SCCR estimator. The latter does not use the information on the covariance structure of the system errors, and therefore not as efficient as the former. The SCCR estimator, in turn, dominates the single equation CCR estimator. As we pointed out above, the SCCR estimation utilizes the system information on the presence of unit roots. It therefore uses some, if not all, system information. This makes even the SCCR estimator unambiguously more efficient than the single equation CCR, which is based solely on the information in each equation.

Corollary 5 *We have*

$$\text{avar}\left(\hat{\beta}_{\text{SUCCR}}\right) \leq \text{avar}\left(\hat{\beta}_{\text{SCCR}}\right) \leq \text{avar}\left(\hat{\beta}_{\text{CCR}}\right)$$

in positive definite sense.

To see the inequalities in Corollary 5 more closely, let us consider a system consisting of two equations with the specification of $\{x_{it}\}$ for $i = 1, 2$ given by M(a) - M(d). Define V_1 and V_2 to be the block diagonal elements of V, and $\Omega_* = (\omega^*_{ij})$ for $i, j = 1, 2$. On the comparison between the SUCCR and SCCR estimators, the standard theory on the system GLS versus OLS applies. The SUCCR and SCCR estimators have the same asymptotic distribution when

$$\mathcal{R}(V\Omega_*) = \mathcal{R}(V) \quad a.s.$$

where and subsequently $\mathcal{R}(M)$ signifies the subspace of $L^2[0,1]$ spanned by the functions in the rows of M. This is completely analogous to the well known Kruskal's theorem on the equivalence of GLS and OLS.

The SUCCR estimator is expected to be strictly more efficient than the SCCR estimator, in general, when

$$\mathcal{R}(V_1) \neq \mathcal{R}(V_2) \quad \text{and} \quad \omega^*_{12} \neq 0$$

i.e., when the SUCCR errors are correlated across the equations in the longrun (or equivalently, when the errors in the original system have nonzero longrun correlation conditional on the differenced stochastic regressors), and the regressors in the two equations have non-overlapping trends, stochastic or deterministic, asymptotically. When the two regressors $\{x_{1t}\}$ and $\{x_{2t}\}$ of the same dimension are cointegrated in a one-to-one fashion (i.e., there are as many cointegrations as the number of regressors), we have in particular that $\mathcal{R}(V_1) = \mathcal{R}(V_2)$ *a.s.* No improvement by the SUCCR estimator is expected in this case, over the SCCR estimator.

When the regressors in the two equations are given by different specifications in M(a) - M(d), or of different dimensions, it is obvious that they would have some distinct trends in the limit. In such cases, nonzero correlation in the SUCCR errors alone

ensures the superiority of the SUCCR estimator. Especially, the SUCCR procedure is in general expected to strictly improve the estimator efficiency when a cointegrating regression is considered jointly with a usual stationary regression on purely deterministic regressors specified by M(d). Other more standard stationary regressions with stochastic regressors may also be allowed here, as long as the regressors are strictly exogenous and the condition PT(a) in Section 3 is satisfied.

For the comparison of the SCCR procedure with that of the single equation CCR, we concentrate on the first equation with the parameter β_1. Denote by $\hat{\beta}^1_{\text{SCCR}}$ and $\hat{\beta}^1_{\text{CCR}}$, respectively, the SCCR and the CCR estimators of β_1. We have

$$\begin{aligned}\sqrt{n}D_{1n}T_1(\hat{\beta}^1_{\text{SCCR}} - \beta_1) &\xrightarrow{\mathcal{D}} \int_{V_1} N\left(0, \omega^*_{11}R(V_1)\right)\, dV_1 \\ \sqrt{n}D_{1n}T_1(\hat{\beta}^1_{\text{CCR}} - \beta_1) &\xrightarrow{\mathcal{D}} \int_{V_1} N\left(0, v_{11}R(V_1)\right)\, dV_1\end{aligned}$$

where D_{1n} and T_1 are the first block diagonal elements of D_n and T, respectively, $R(V_1) = (\int_0^1 V_1 V_1')^{-1}$, and v_{11} is the conditional longrun variance of $\{u_{1t}\}$ given $\{\Delta x^0_{1t}\}$. Other notations V_1 and ω^*_{11} are as defined earlier. Obviously,

$$\omega^*_{11} \leq v_{11}$$

since ω^*_{11} is the longrun variance of $\{u_{1t}\}$ conditional on $\{\Delta x^0_t\}$.

To see exactly when the SCCR estimator is strictly better than the single equation CCR estimator, we let v_{12} be the longrun covariance between $\{u_{1t}\}$ and $\{\Delta x^0_{2t}\}$ conditional on $\{\Delta x^0_{1t}\}$ and $v_{21} = v'_{12}$. Similarly, Υ_{22} is defined to be the conditional longrun variance of $\{\Delta x^0_{2t}\}$ given $\{\Delta x^0_{1t}\}$. It follows then that

$$\omega^*_{11} = v_{11} - v_{12}\Upsilon^{-1}_{22}v_{21}$$

Now it is clear that the unit roots in the second equations have a potential to increase the efficiency of an estimate for β_1, when and only when $\{\Delta x^0_{2t}\}$ is correlated in the longrun with $\{u_{1t}\}$, conditional on $\{\Delta x^0_{1t}\}$. As can be easily checked, the conditional covariance is zero when $\{x^0_{1t}\}$ and $\{x^0_{2t}\}$ are cointegrated. No improvement of the SCCR estimator upon the single equation CCR is expected in such a case. Unlike in the comparison of SUCCR and SCCR, stationary regressions cannot improve the

SCCR estimator over the single equation CCR estimator, because they include no unit roots.

In sum, the efficiency gain is made by the SUCCR procedure in two steps. In the first step, it improves upon the single equation methods by using the system information on the presence of unit roots in individual series. This is done through the construction of the SUCCR transformation. In the second step, the SUCCR estimation subsequently gains efficiency by using the system GLS method, based on the longrun covariance matrix of the system errors (which are rightfully rebuilt by making necessary transformations). Therefore, the SUCCR procedure fully utilizes the system information, including both the unit roots and the error covariance structure of the system.

Due to the mixed normality of the limiting distributions of the least squares estimators, the usual chi-square tests on the coefficient β in the SUCCR system (22) are possible. To show this more precisely, we consider a general hypothesis of the form

$$H_0 : \quad \varphi(\beta) = 0 \tag{27}$$

where the function φ is assumed to be continuously differentiable with the first derivative Φ evaluated at the true value of β. Assume that there are q restrictions under H_0.

For the usual chi-square test relying on the SUCCR estimator, it is straightforward to show using the result and the proof of Corollary 3 that

Corollary 6 *Under the null hypothesis* (27), *we have*

$$\varphi(\hat{\beta}_{\text{SUCCR}})' \left(\Phi(X_*'(\Omega_*^{-1} \otimes I)X_*)^{-1}\Phi'\right)^{-1} \varphi(\hat{\beta}_{\text{SUCCR}}) \xrightarrow{\mathcal{D}} \chi_q^2$$

Moreover,

$$\varphi(\hat{\beta}_{\text{SCCR}})' \left(\Phi(X_*'X_*)^{-1}X_*'(\Omega_*^{-1} \otimes I)X_*(X_*'X_*)^{-1}\Phi'\right)^{-1} \varphi(\hat{\beta}_{\text{SCCR}}) \xrightarrow{\mathcal{D}} \chi_q^2$$

The tests on β can therefore be done in the SUCCR system precisely as in the standard SUR, except that the longrun variance Ω_* of the errors must be used. This

modification of the weighting matrix is necessary in our general context that allows for serial correlations in the errors.

5. Feasible SUCCR

For the practical application of the SUCCR method, the transformation in (20), as well as the longrun variance Ω_* of the SUCCR error $\{u_t^*\}$, must be consistently estimated. The consistent estimation of the transformations for each equation can essentially be done, following Park (1991) and Park and Ogaki (1991). The construction of the SUCCR system, however, can be quite elusive for the SUCR system consisting of the equations generated by a mixture of M(a) - M(d). In this section, we will explain in detail how to construct the SUCCR transformation in applications.

The transformation in (20) may be rewritten in matrix form as

$$\begin{aligned} y_* &= y - \mathrm{vec}(W_2\Omega_{22}^{-1}\Omega_{21}) - (I \otimes W\Sigma^{-1})\Lambda\beta \\ X_* &= X - (I \otimes W\Sigma^{-1})\Lambda \end{aligned}$$

where $\Lambda = \mathrm{diag}(\Lambda_1, \ldots, \Lambda_m)$, $W = (w_1, \ldots, w_n)'$ and W_2 is defined as the second column block of W, i.e., $W = (W_1, W_2)$ with the partition made according to $w_t = (u_t', \Delta x_t^{0\prime})'$. The operator 'vec($\cdot$)' vectorizes a matrix by stacking columns, and 'diag($\cdot$)' denotes a block diagonal matrix as before. Other notations are defined earlier.

First, there are 'natural' consistent estimates of β and $\{w_t\}$, which can be obtained by the single equation CCR for each equation in the SUCR system (1), and the OLS regressions based upon M(a) - M(c). From the CCR estimate $\hat{\beta}^i_{\mathrm{CCR}}$ for β_i, we may consistently estimate $\{u_{it}\}$ by

$$\hat{u}_{it} = y_{it} - x_{it}'\hat{\beta}^i_{\mathrm{CCR}}$$

Any consistent estimate for β_i, such as the OLS estimate, can also be used. More efficient CCR procedure seems, however, desirable to improve the finite sample performance of the SUCCR estimator. A consistent estimate for $\{\Delta x_t^0\}$ can be obtained

from $\widehat{\Delta x}_{it}^{0} = \Delta x_{it}$ for the equations given by M(a), or from the OLS residual in the regression

$$\Delta x_{it}, \Delta q_{it} \;=\; \hat{\pi}_i \Delta p_{it} + \widehat{\Delta x}_{it}^{0}$$

for the equations specified by M(b) and M(c). Now we may estimate $\{w_t\}$ by

$$\hat{w}_t = (\hat{u}_t', \widehat{\Delta x}_t^{0\prime})'$$

Second, the rest of the parameters are estimated from $\{\hat{w}_t\}$. Before we explain how this can be done, it is necessary to introduce some additional notation. Define

$$\Gamma = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n-1} \sum_{t=k+1}^{n} E(w_t w_{t-k}') = (\Gamma_1, \Gamma_2)$$

where Γ_2 is a submatrix of Γ consisting of columns corresponding to $\{\Delta x_t^0\}$ in $w_t = (u_t', \Delta x_t^{0\prime})'$. Define Σ_2 from Σ in (8) similarly. Let

$$\Lambda^0 = \Sigma_2 + \Gamma_2$$

Notice that $\Omega = \Sigma + \Gamma + \Gamma'$.

A consistent estimate of Λ can easily be obtained from that of Λ^0 by transforming it into a diagonal matrix, and augmenting zeros for the purely deterministic regressors $\{p_{it}\}$ and $\{x_{it}\}$ in the equations given respectively as M(c) and M(d). Let Λ_i^0 be the matrix consisting of the columns of Λ^0 corresponding to $\{\Delta x_{it}^0\}$. Then the i-th block diagonal of Λ is Λ_i^0 for the equation with M(a) or M(b), and the matrix of zeros juxtaposed by Λ_i^0 for the equation with M(c), and finally the matrix of zeros for the equation with M(d). The zero matrices for the equations given by M(c) and M(d) have the number of columns equal to the number of deterministic regressors.

For the consistent estimation of the parameters Ω and Λ^0, it suffices to estimate Σ and Γ consistently. We may consistently estimate Σ simply by

$$\hat{\Sigma} = \frac{1}{n} \sum_{t=1}^{n} \hat{w}_t \hat{w}_t'$$

There are various methods to estimate Γ. It can be estimated nonparametrically as

$$\hat{\Gamma} = \frac{1}{n} \sum_{k\geq 1} c(k) \sum_{t=k+1}^{n} \hat{w}_t \hat{w}_{t-k}'$$

with a chosen weight function $c(k)$. The reader is referred to Andrews (1988) for a detailed discussion on the choice of the weight function.

Parametric estimation is also possible. If $\{w_t\}$ follows a p-th order VAR

$$w_t = \sum_{k=1}^{p} \Phi_k w_{t-k} + e_t$$

where $\{e_t\}$ is white noise, then we may deduce that

$$\Gamma = \left(\mathrm{I} - \sum_{i=1}^{p} \Phi_i\right)^{-1} \sum_{i=0}^{p-1} \sum_{j=i+1}^{p} \Sigma_i' \Phi_j$$

where

$$\Sigma_k = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E(w_t w_{t-k}')$$

as shown in Park and Ogaki (1991). The parameter Γ can therefore be consistently estimated from the estimated VAR coefficients $\hat{\Phi}_k$, and $\hat{\Sigma}_k = \sum_{t=k+1}^{n} \hat{w}_t \hat{w}_{t-k}'/n$.

The case that $\{w_t\}$ has a VAR structure naturally arises when the cointegrating relationships in (1) are generated by an ECM of finite order. Let $z_t^0 = (y_t^{0\prime}, x_t^{0\prime})'$, where $\{y_t^0\}$ is the stochastic component of $\{y_t\}$ similarly as $\{x_t^0\}$. If $\{z_t^0\}$ is driven by a VAR of order p, $\{w_t\}$ is indeed given precisely as the above p-th order VAR with the restriction that the coefficients in the submatrix of Φ_p corresponding to $\{\Delta x_{t-p}^0\}$ are zero. This follows immediately from the ECM representation of such a system, as shown in Park and Ogaki (1991).

Finally, a consistent estimate for the longrun variance Ω_* of the SUCCR errors, which we need to compute the SUCCR estimator and to test on the coefficient β, can be obtained either directly from the fitted SUCCR errors $\{\hat{u}_t^*\}$ or from a consistent estimate of Ω using the relationship in (25).

6. Monte Carlo Simulation

Monte Carlo results for the finite sample performance of the SUCCR estimator are summarized in this section. In the simulation, we paid a particular attention to the relative performance of the SUCCR estimator compared with the SCCR estimator

and the single equation CCR estimator. The comparison of the SUCCR estimator with the SCCR estimator would reveal how much gain in efficiency we may expect in finite samples through the application of the GLS method. This is the gain in efficiency coming from the utilization of the covariance structure of the system errors. In contrast, the comparison of the SCCR estimator and the single equation CCR estimator would tell us how important in finite samples it is to use the system information on the presence of the unit roots.

The simulations reported in this paper are based on a SUCR system consisting of two equations with $\beta_1 = \beta_2 = 1$. The DGP for each equation is given by the first order ECM specified as

$$\begin{pmatrix} \Delta y_{it} \\ \Delta x_{it} \end{pmatrix} = \begin{pmatrix} \alpha_{1i} \\ \alpha_{2i} \end{pmatrix} (y_{i,t-1} - x_{i,t-1}) + \begin{pmatrix} \epsilon^1_{it} \\ \epsilon^2_{it} \end{pmatrix}$$

for $i = 1, 2$. The covariance matrix of $\epsilon_t = (\epsilon^1_{1t}, \epsilon^1_{2t}, \epsilon^2_{1t}, \epsilon^2_{2t})$ is denoted by Υ which is given by

$$\Upsilon = \begin{pmatrix} 1 & \rho_1 & \rho_0 & 0 \\ \rho_1 & 1 & 0 & \rho_0 \\ \rho_0 & 0 & 1 & \rho_2 \\ 0 & \rho_0 & \rho_2 & 1 \end{pmatrix}$$

The error correction coefficients α_{11} and α_{12} are set -0.2, and $\alpha_{21} = \alpha_{22} = 0.1, 0.2, 0.3$ and 0.4. The correlation coefficient between the errors of the equations for Δy_{it} and Δx_{it}, signified by ρ_0, is set $0.1, 0.3$ and 0.5. The cross-equation correlations in the errors of Δy_{1t} and Δy_{2t} and those of Δx_{1t} and Δx_{2t} are denoted by ρ_1 and ρ_2, respectively. The values of (ρ_1, ρ_2) used in the simulation were $(0.5, 0.2), (0.2, -0.5)$ and $(-0.2, 0.6)$. The values of ρ_1 and ρ_2 are irrelevant for the single equation estimators. They are, however, very important for the system estimation, since they critically affect the longrun covariances across the equations of the regressors and the regression errors.

To concentrate on the comparisons of the single equation CCR and the system SCCR and SUCCR estimators, we assume that the underlying data generating process is known to be VAR of first order. The CCR and SUCCR transformations are therefore estimated parametrically, as explained in the previous section. For the fi-

nite sample comparisons of the CCR and other efficient procedures such as the exact ML method, the reader is referred to Park and Ogaki (1991) and the references cited there. The effect of the parametric and nonparametric estimations of the CCR transformation on the finite sample efficiency is also investigated extensively there. To avoid the dependency of the simulation results on initialization, the constant term was included in all the estimation procedures.

The simulation results are tabulated in Tables 1 and 2, respectively for the samples of size 100 and 300. The biases, MSE's and the theoretical variances are given in the tables. The theoretical variances of the estimators in each model are computed from the asymptotic theory presented in the previous section. To make the results in Table 1 and Table 2 more directly comparable, we adjusted the MSE's and the theoretical variances for the samples of size 300 by multiplying by 9. The theoretical variances in Table 1 and Table 2 are therefore identical. For the same reason, the biases for the sample size 300 were multiplied by 3.

The relative performance of the estimators in finite samples largely coincides with what is expected from the asymptotic theory. The system SUCCR and SCCR estimators improve upon the single equation CCR estimator in finite samples, often significantly. As the theory suggests, the SUCCR estimator has the smallest MSE's in most cases. Likewise, the SCCR estimator has smaller MSE's than the single equation CCR estimator in a majority of cases. The advantage of the system estimation seems to be evident and practically important. It appears that the system estimation is clearly preferred even for the samples of moderate size. The system SCCR and SUCCR methods also significantly reduce the finite sample bias, in many cases, over the single equation CCR procedure.

For the samples of size 100, however, the reduction in variance through the system estimation does not seem to be as big as the theory suggests. The actual finite sample variances of the both system SUCCR and SCCR estimators were farther away from the theoretical variances, than those of the single equation CCR estimator. When the differences in the theoretical variances of system and single equation estimators are only marginal, the single equation CCR indeed performed slightly better. Of the

two system estimators, the divergence of the actual variance from the asymptotic theoretical variance appears to be larger for the SUCCR estimator. Consequently, in some of the models for which the cross correlations of the SUCCR errors are small and the system GLS is expected to have no substantial improvement over the OLS, the SCCR estimators had MSE's smaller than those of the SUCCR estimators.

Virtually all the discrepancies between the asymptotic theory and the actual performance in the comparison of the estimators disappear, when the sample size was increased to 300. In the models that we looked at for the simulation, the actual variances of the single equation CCR were still somewhat closer to the theoretical variances than those of the system estimators. The finite sample MSE's continue to better approximate the asymptotic variances for the SCCR estimators, than for the SUCCR estimators. Yet, the differences were not significant enough to reverse the comparison among the estimators. They were, in particular, much smaller in magnitude than those for the samples of size 100. As a result, the SCCR estimator outperformed the single equation CCR estimator, and the SUCCR estimator had MSE's smaller than those of the SCCR estimator, with only few exceptions. All the exceptional cases were when the asymptotic theories do not predict any meaningful improvements.

Finally, we observed that the GLS procedure in the SUCCR system sometimes, though not very often, yields the estimates that are far away from the true parameter value, when the sample size is 100. The SUCCR estimator, in relatively small samples, seems unstable and can be bad. The OLS estimator in the SUCCR system, the SCCR estimator, does not seem to have this problem. Though having yielded larger MSE's than the SUCCR estimator in most cases, the SCCR estimator was quite stable across various data generating processes. In small samples, the SCCR estimator appears to be a reasonable alternative to the SUCCR estimator. The instability of the SUCCR estimator quickly vanished as we increased the sample size in the simulation. When the sample size was 300, the problem was non-existent.

7. Concluding Remarks

The SUR system has been, and no doubt will continuously be, looked at to study various economic models. Naturally, we may postulate each equation in a SUR system as representing a longrun cointegrating relationship, if the variables in the system are individually modelled as possessing unit roots. We have in this paper developed a method of inference in cointegrated models formulated as the SUR system, which we called the SUCR in the paper. Our framework is broad, and incorporates deterministic as well as stochastic cointegration. It also allows for a system consisting of a mixture of cointegrating regressions and the usual stationary regressions. The proposed procedure yields an efficient parameter estimate, and the usual chi-square tests on coefficient restrictions. In the context of SUCR, all the existing methods are only applicable equation by equation. Such an application would not only make it impossible to perform tests on cross equation restrictions, but also yield sub-optimal estimates for the parameters in the model.

The purpose of this paper is to provide a method of inference in a model specified as a SUCR. The specification tests to check the adequacy of such a model are, therefore, only briefly mentioned. The existing tests of cointegration and non-cointegration can be applied to check the adequacy of the specification of each cointegrating relationship. The system-wise test is also possible, directly using the variable addition approach by Park, Ouliaris and Choi (1988) and Park (1990a). This will be reported elsewhere. One final remark on the use of testing for cointegration as a specification test for a SUCR: an appropriately formulated SUCR implicitly defines the number of cointegrations for a given set of variables. The tests for the number of cointegration in the set of variables alone, however, is not sufficient to justify any particular specification of a SUCR. The formulation of a SUCR system requires not only that a certain number of cointegrations be present in the model, but also that they be given in a specific, identified form.

Appendix: Mathematical Proofs

Proof of Theorem 1. Let

$$z_{it} = x_{it}^0, \begin{pmatrix} p_{it} \\ H_{2i}' x_{it}^0 \end{pmatrix}, \begin{pmatrix} p_{it} \\ x_{it}^0 \end{pmatrix}, p_{it}$$

for models M(a) – M(d), respectively. Define a matrix Z from $\{z_{it}\}$, similarly as X from $\{x_{it}\}$. It follows directly from Lemma (A1) of Park (1991) that

$$\begin{aligned} \tfrac{1}{n} D_n^{-2} Z'Z &\xrightarrow{\mathcal{D}} \int_0^1 VV' \\ \tfrac{1}{n} D_n^{-2} Z' \hat{\Sigma}_0^{-1} Z &\xrightarrow{\mathcal{D}} \int_0^1 V \Sigma_0^{-1} V' \end{aligned}$$

and

$$\begin{aligned} \tfrac{1}{\sqrt{n}} D_n^{-1} Z'u &\xrightarrow{\mathcal{D}} \int_0^1 V\, dU + \mu \\ \tfrac{1}{\sqrt{n}} D_n^{-1} Z' \hat{\Sigma}_0^{-1} u &\xrightarrow{\mathcal{D}} \int_0^1 V \Sigma_0^{-1}\, dU + \nu \end{aligned}$$

To get the stated results, write

$$z_{it} = T_i' x_{it}$$

where T_i is defined in (16), and

$$\begin{aligned} y_{it} &= x_{it}'\beta_i + u_{it} \\ &= z_{it}'\delta_i \end{aligned}$$

where

$$\delta_i = T_i^{-1}\beta_i$$

Proof of Lemma 2. Under the condition PT(a), we have

$$\lambda_{ij} = 0$$

for all $i, j = 1, \ldots, m$, and therefore,

$$\mu = \nu = 0$$

Furthermore, since $\Omega_{12} = 0$, the limit Brownian motions B_1 and B_2 are independent, which implies in turn that V is independent of U.

We have under the condition PT(b),

$$E(dU\, dU') = \Sigma_0$$

Therefore,

$$\begin{aligned} \left(\int_0^1 V\Sigma_0^{-1}V'\right)^{-1} \int_0^1 V\Sigma_0^{-1}\, dU \,\Big|_V &\sim N\left(0, P(V,\Sigma_0)\right) \\ \left(\int_0^1 VV'\right)^{-1} \int_0^1 V\, dU \,\Big|_V &\sim N\left(0, Q(V,\Sigma_0)\right) \end{aligned}$$

where $\cdot\,|_V$ denotes the conditional distribution given V. This was to be shown.

Proof of Corollary 3. The stated result follows immediately, since

$$P(V,\Sigma_0) \;\le\; Q(V,\Sigma_0) \quad a.s.$$

and

$$\begin{aligned} \text{var} \int_V N(0, P(V,\Sigma_0))\, dV &= E\, P(V,\Sigma_0) \\ \text{var} \int_V N(0, Q(V,\Sigma_0))\, dV &= E\, Q(V,\Sigma_0) \end{aligned}$$

Proof of Theorem 4. Define $\{x_{it}^{*0}\}$ to be the stochastic components of the regressors in the i-th equation of the SUCCR (22), and

$$\lambda_{ij}^* = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E(x_{it}^{*0} u_{jt}^*)$$

similarly as λ_{ij} in (17) for the original model (1). From the orthogonality of the regressors and the regression errors in the SUCCR system noted in (23), we have

$$\lambda_{ij}^* = 0$$

for all $i, j = 1, \ldots, m$. Moreover, the SUCCR errors $\{u_t^*\}$ are asymptotically independent of the stochastic regressors $\{x_t^{*0}\}$, $x_t^{*0} = (x_{1t}^{*0\prime}, \ldots, x_{mt}^{*0\prime})'$, or $\{x_t^0\}$, as is easily seen from (24). The stated results, therefore, follow exactly as in the proof of Lemma 2.

Proof of Corollary 5. Exactly as in the proof of Corollary 3,

$$P(V,\Omega_*) \leq Q(V,\Omega_*) \quad a.s.$$

and the first inequality follows directly.

The single equation CCR error $\{u_{it}^+\}$, say, is given from Park (1991) by

$$u_{it}^+ = u_{it} - \omega_{12}^{ii}\Omega_{22}^{ii^{-1}}\triangle x_{it}^0$$

where Ω_{22}^{ii} is the longrun variance of $\{\triangle x_{it}^0\}$, and ω_{12}^{ii} is the longrun covariance between $\{u_{it}\}$ and $\{\triangle x_{it}^0\}$. Consequently,

$$\sqrt{n}D_nT(\hat{\beta}_{\text{CCR}} - \beta) \xrightarrow{\mathcal{D}} \int_V N\left(0, P(V,\Omega_+)\right)\, dV$$

where Ω_+ is the longrun variance of the single equation CCR error $\{u_t^+\}$, $u_t^+ = (u_{1t}^+, \ldots, u_{mt}^+)'$. Clearly,

$$\Omega_* \leq \Omega_+$$

and

$$Q(V,\Omega_*) \leq Q(V,\Omega_+) \quad a.s.$$

as was to be shown.

Proof of Corollary 6. The stated results follow easily from Theorem 4 and Corollary 4.2 of Park (1991).

References

[1] Andrews, D.W.K. (1990). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," Cowles Foundation Discussion Paper No. 877R, Yale University, forthcoming in *Econometrica.*

[2] Engle, R.F. and C.W.J. Granger (1987). "Co-integration and error correction: Representation, estimation and testing," *Econometrica*, 55, 251-276.

[3] Johansen, S. (1988). "Statistical analysis of cointegrating vectors," *Journal of Economic Dynamics and Control*, 12, 231-254

[4] Johansen, S. (1989). "Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models," University of Copenhagen, Institute of Mathematical Statistics Preprint No. 3.

[5] Ogaki, M. (1990) "Engle's law and cointegration," Rochester Center for Economic Research Working Paper No. 228, University of Rochester.

[6] Ogaki, M. and J.Y. Park (1990) "A cointegrating approach to estimating preference parameter," Rochester Center for Economic Research Working Paper No. 236, University of Rochester.

[7] Park, J.Y. (1990a). "Testing for unit roots and cointegration by variable addition," in Rhodes, G.F. and T.B. Fomby ed., *Advances in Econometrics*, 8, 107-133.

[8] Park, J.Y. (1990b). "Maximum likelihood estimation of simultaneous cointegrated models," Economic Institute Memo No. 1990-18, Aarhus University.

[9] Park, J.Y. (1991). "Canonical cointegrating regressions," CAE Working Paper No. 88-29R, Cornell University, forthcoming in *Econometrica.*

[10] Park, J.Y. and M. Ogaki (1991) "Inference in cointegrated models using VAR prewhitening to estimate shortrun dynamics," mimeograph, Cornell University and University of Rochester.

[11] Park, J.Y. and J. Sung (1989) "An empirical investigation of employment, wages and output: A comparative study of US and Japan," CAE Working Paper No. 89-26, Cornell University.

[12] Park, J.Y., S. Ouliaris and B. Choi (1988). "Spurious regressions and tests for cointegration," CAE Working Paper No. 88-07, Cornell University.

[13] Park, J.Y. and P.C.B. Phillips (1988). "Statistical inference in regressions with integrated processes: Part 1," *Econometric Theory*, 4, 468-497.

[14] Park, J.Y. and P.C.B. Phillips (1989). "Statistical inference in regressions with integrated processes: Part 2," *Econometric Theory*, 5, 95-131.

[15] Phillips, P.C.B. (1988). "Spectral regression for cointegrated time series," Cowles Foundation Discussion Paper No. 872, Yale University.

[16] Phillips, P.C.B. (1991). "Optimal inference in cointegrated systems," *Econometrica*, 59, 283-306.

[17] Phillips, P.C.B. and S.N. Durlauf (1986). "Multiple time series with integrated variables," *Review of Economic Studies*, 53, 473-496.

[18] Phillips, P.C.B. and B.E. Hansen (1990). "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies*, 57, 99-126.

[19] Phillips, P.C.B. and S. Ouliaris (1990). "Asymptotic properties of residual based tests for cointegration," *Econometrica*, 58, 165-193.

[20] Phillips, P.C.B. and J.Y. Park (1988). "Asymptotic equivalence of OLS and GLS in regressions with integrated regressors," *Journal of the American Statistical Association*, 83, 111-115.

[21] Stock, J.H. and M.W. Watson (1988). "Testing for common trends," *Journal of the American Statistical Association*, 83, 1097-1107.

[22] Stock, J.H. and M.W. Watson (1991). "A simple estimator of cointegrating vectors in higher order integrated systems," mimeograph, University of California, Berkeley and Northwestern University.

[23] West, K.D. (1988). "Asymptotic normality, when regressors have a unit root," *Econometrica*, 56, 1397-1417.

[24] Zellner, A. (1962). "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," *Journal of the American Statistical Association*, 57, 348-368.

Table 1

Finite Sample Bias, MSE and Asymptotic Variance: $n = 100$

Single Equation Estimators

		OLS		CCR		
α_2	ρ_0	Bias	MSE	Bias	MSE	Avar
0.1	0.1	−0.2249	0.0857	−0.0752	0.0387	0.0330
	0.3	−0.1593	0.0482	−0.0482	0.0255	0.0230
	0.5	−0.1046	0.0246	−0.0290	0.0160	0.0149
0.2	0.1	−0.2112	0.0727	−0.0507	0.0251	0.0221
	0.3	−0.1463	0.0387	−0.0321	0.0162	0.0145
	0.5	−0.0943	0.0187	−0.0194	0.0099	0.0090
0.3	0.1	−0.1897	0.0579	−0.0334	0.0155	0.0133
	0.3	−0.1324	0.0308	−0.0219	0.0102	0.0089
	0.5	−0.0858	0.0148	−0.0137	0.0063	0.0056
0.4	0.1	−0.1724	0.0472	−0.0228	0.0097	0.0082
	0.3	−0.1220	0.0255	−0.0157	0.0066	0.0058
	0.5	−0.0800	0.0123	−0.0102	0.0042	0.0037

System Estimators

(a): $\rho_1 = 0.5$, $\rho_2 = 0.2$

		SCCR			SUCCR		
α_2	ρ_0	Bias	MSE	Avar	Bias	MSE	Avar
0.1	0.1	−0.0206	0.0389	0.0326	−0.0236	0.0348	0.0265
	0.3	−0.0115	0.0278	0.0228	−0.0148	0.0219	0.0167
	0.5	−0.0059	0.0179	0.0148	−0.0116	0.0105	0.0073
0.2	0.1	−0.0089	0.0254	0.0215	−0.0097	0.0242	0.0187
	0.3	−0.0050	0.0168	0.0141	−0.0072	0.0135	0.0110
	0.5	−0.0026	0.0104	0.0087	−0.0066	0.0061	0.0046
0.3	0.1	−0.0041	0.0155	0.0129	−0.0055	0.0140	0.0116
	0.3	−0.0026	0.0102	0.0086	−0.0041	0.0084	0.0070
	0.5	−0.0014	0.0063	0.0053	−0.0041	0.0038	0.0030
0.4	0.1	−0.0024	0.0099	0.0080	−0.0034	0.0089	0.0074
	0.3	−0.0013	0.0065	0.0055	−0.0025	0.0054	0.0046
	0.5	−0.0008	0.0041	0.0035	−0.0027	0.0025	0.0021

Table 1: *Continued*

(b): $\rho_1 = 0.2$, $\rho_2 = -0.5$

		SCCR			SUCCR		
α_2	ρ_0	Bias	MSE	Avar	Bias	MSE	Avar
0.1	0.1	–0.0193	0.0354	0.0296	–0.0186	0.0374	0.0296
	0.3	–0.0110	0.0241	0.0200	–0.0100	0.0246	0.0198
	0.5	–0.0060	0.0146	0.0122	–0.0044	0.0145	0.0116
0.2	0.1	–0.0082	0.0219	0.0193	–0.0061	0.0223	0.0186
	0.3	–0.0049	0.0144	0.0126	–0.0033	0.0141	0.0118
	0.5	–0.0030	0.0086	0.0075	–0.0013	0.0080	0.0066
0.3	0.1	–0.0037	0.0135	0.0119	–0.0023	0.0130	0.0108
	0.3	–0.0024	0.0090	0.0079	–0.0011	0.0084	0.0070
	0.5	–0.0017	0.0055	0.0048	–0.0003	0.0048	0.0040
0.4	0.1	–0.0017	0.0085	0.0076	–0.0006	0.0080	0.0066
	0.3	–0.0012	0.0059	0.0052	–0.0003	0.0054	0.0045
	0.5	–0.0010	0.0037	0.0033	0.0000	0.0031	0.0026

(c): $\rho_1 = -0.2$, $\rho_2 = 0.6$

		SCCR			SUCCR		
α_2	ρ_0	Bias	MSE	Avar	Bias	MSE	Avar
0.1	0.1	–0.0184	0.0325	0.0283	–0.0176	0.0330	0.0282
	0.3	–0.0104	0.0213	0.0187	–0.0013	0.3078	0.0183
	0.5	–0.0060	0.0127	0.0111	–0.0030	0.0124	0.0097
0.2	0.1	–0.0084	0.0204	0.0184	–0.0060	0.0200	0.0170
	0.3	–0.0052	0.0133	0.0119	–0.0031	0.0124	0.0105
	0.5	–0.0034	0.0078	0.0070	–0.0011	0.0064	0.0052
0.3	0.1	–0.0040	0.0129	0.0114	–0.0022	0.0116	0.0097
	0.3	–0.0028	0.0085	0.0076	–0.0013	0.0073	0.0061
	0.5	–0.0021	0.0051	0.0046	–0.0006	0.0038	0.0031
0.4	0.1	–0.0019	0.0083	0.0074	–0.0005	0.0071	0.0059
	0.3	–0.0015	0.0057	0.0051	–0.0004	0.0046	0.0039
	0.5	–0.0014	0.0036	0.0032	–0.0003	0.0024	0.0020

Note: The simulations were based on the samples generated by the random number generator built in GAUSS-386. All the computations were done using programs written in GAUSS. The number of iteration is 5000.

Table 2

Finite Sample Bias, MSE and Asymptotic Variance: $n = 300$

Single Equation Estimators

		OLS		CCR		
α_2	ρ_0	Bias	MSE	Bias	MSE	Avar
0.1	0.1	−0.2676	0.1315	−0.0400	0.0359	0.0330
	0.3	−0.1856	0.0704	−0.0242	0.0243	0.0230
	0.5	−0.1198	0.0343	−0.0141	0.0155	0.0149
0.2	0.1	−0.2501	0.1110	−0.0251	0.0236	0.0221
	0.3	−0.1685	0.0551	−0.0152	0.0153	0.0145
	0.5	−0.1063	0.0251	−0.0090	0.0094	0.0090
0.3	0.1	−0.2205	0.0837	−0.0150	0.0138	0.0133
	0.3	−0.1498	0.0417	−0.0096	0.0093	0.0089
	0.5	−0.0950	0.0188	−0.0060	0.0058	0.0056
0.4	0.1	−0.1973	0.0652	−0.0095	0.0085	0.0082
	0.3	−0.1361	0.0332	−0.0065	0.0059	0.0058
	0.5	−0.0875	0.0151	−0.0043	0.0038	0.0037

System Estimators

(a): $\rho_1 = 0.5$, $\rho_2 = 0.2$

		SCCR			SUCCR		
α_2	ρ_0	Bias	MSE	Avar	Bias	MSE	Avar
0.1	0.1	−0.0025	0.0347	0.0326	−0.0026	0.0290	0.0265
	0.3	−0.0016	0.0244	0.0228	−0.0026	0.0183	0.0167
	0.5	−0.0012	0.0159	0.0148	−0.0040	0.0082	0.0073
0.2	0.1	−0.0013	0.0230	0.0215	−0.0023	0.0202	0.0187
	0.3	−0.0010	0.0151	0.0141	−0.0022	0.0119	0.0110
	0.5	−0.0008	0.0093	0.0087	−0.0028	0.0051	0.0046
0.3	0.1	−0.0006	0.0136	0.0129	−0.0015	0.0124	0.0116
	0.3	−0.0006	0.0090	0.0086	−0.0017	0.0075	0.0070
	0.5	−0.0006	0.0056	0.0053	−0.0019	0.0032	0.0030
0.4	0.1	−0.0003	0.0084	0.0080	−0.0011	0.0078	0.0074
	0.3	−0.0004	0.0057	0.0055	−0.0013	0.0049	0.0046
	0.5	−0.0004	0.0036	0.0035	−0.0014	0.0022	0.0021

Table 2: *Continued*

(b): $\rho_1 = 0.2$, $\rho_2 = -0.5$

		SCCR			SUCCR		
α_2	ρ_0	Bias	MSE	Avar	Bias	MSE	Avar
0.1	0.1	–0.0018	0.0311	0.0296	–0011	0.0312	0.0296
	0.3	–0.0010	0.0210	0.0200	–0004	0.0209	0.0198
	0.5	–0.0009	0.0128	0.0122	–0000	0.0123	0.0116
0.2	0.1	–0.0010	0.0201	0.0193	0000	0.0195	0.0186
	0.3	–0.0009	0.0131	0.0126	0001	0.0124	0.0118
	0.5	–0.0009	0.0078	0.0075	0001	0.0069	0.0066
0.3	0.1	–0.0005	0.0123	0.0119	0006	0.0113	0.0108
	0.3	–0.0006	0.0082	0.0079	0004	0.0073	0.0070
	0.5	–0.0007	0.0050	0.0048	0003	0.0041	0.0040
0.4	0.1	–0.0002	0.0078	0.0076	0007	0.0069	0.0066
	0.3	–0.0004	0.0054	0.0052	0005	0.0046	0.0045
	0.5	–0.0006	0.0034	0.0033	0003	0.0027	0.0026

(c): $\rho_1 = -0.2$, $\rho_2 = 0.6$

		SCCR			SUCCR		
α_2	ρ_0	Bias	MSE	Avar	Bias	MSE	Avar
0.1	0.1	–0.0026	0.0298	0.0283	–0.0015	0.0301	0.0282
	0.3	–0.0020	0.0198	0.0187	–0.0009	0.0196	0.0183
	0.5	–0.0017	0.0117	0.0111	–0.0002	0.0106	0.0097
0.2	0.1	–0.0016	0.0191	0.0184	–0.0008	0.0181	0.0170
	0.3	–0.0013	0.0124	0.0119	–0.0007	0.0112	0.0105
	0.5	–0.0011	0.0073	0.0070	–0.0003	0.0057	0.0052
0.3	0.1	–0.0004	0.0118	0.0114	–0.0003	0.0102	0.0097
	0.3	–0.0006	0.0078	0.0076	–0.0004	0.0065	0.0061
	0.5	–0.0006	0.0047	0.0046	–0.0003	0.0033	0.0031
0.4	0.1	–0.0000	0.0076	0.0074	–0.0003	0.0062	0.0059
	0.3	–0.0002	0.0052	0.0051	–0.0004	0.0040	0.0039
	0.5	–0.0004	0.0032	0.0032	–0.0004	0.0021	0.0020

Note: The actual biases and MSE's were multiplied, respectively, by three and nine to obtain the numbers reported here. This makes the results in this table directly comparable to those reported in Table 1 for the samples of size 100. The asymptotic variances in two tables are identical. The simulations were done exactly as in Table 1.