3-2002

Simple Tests for Reduced Rank in Multivariate Regression

Scott Gilbert
Southern Illinois University Carbondale

Petr Zemcik
Southern Illinois University Carbondale

Simple Tests for Reduced Rank in Multivariate Regression

Scott Gilbert* and Petr Zemčík
Department of Economics
Southern Illinois University at Carbondale

March 2002

Abstract

The present work proposes tests for reduced rank in multivariate regression coefficient matrices, under rather general conditions. A heuristic approach is to first estimate the regressions via standard methods, then compare the coefficient matrix rows (or columns) to assess their redundancy. A formal version of this approach utilizes the distance between an unrestricted coefficient matrix estimate and an estimate restricted by reduced rank. Two distance minimization problems emerge, based on equivalent formulations of the null hypothesis. For each method we derive estimators and tests, and their asymptotic distributions. We examine test performance in simulation, and give some numerical examples.

Keywords: Reduced rank, multivariate regression, asymptotic theory, simulation.

JEL Code: C3 (Econometric Methods: Multiple/Simultaneous Equation Models)

* *Correspondence address:* Department of Economics, Mailcode 4515, Southern Illinois University at Carbondale, Carbondale, IL 62901-4515, Tel:(618) 453-5065, Fax: (618) 453-2717, E-mail: gilberts@siu.edu.

1. Introduction

A major tool of econometrics is the multivariate linear regression model, with two or more equations, each specifying a relationship between a variable of interest and some explanatory variables. The coefficients of the regression model are parameters that form a matrix composed of two or more row vectors, each vector describing a different regression equation. Inference for multivariate regression models typically centers on the coefficient matrix and hypotheses that imply restrictions on this matrix.

Most econometric testing of multivariate regressions has focused on the validity of hypotheses expressible as linear restrictions on the coefficient matrix; however, economic theory is often consistent with more complex, non-linear restrictions. Among such restrictions are those concerning the rank of (some part of) the coefficient matrix β, e.g. the number of linearly independent matrix rows or, equivalently, the number of linearly independent columns in β. With g rows and K columns, the rank of β is no more than the minimum $\min(g, K)$, and the matrix has *reduced* rank (alternatively *short* rank, as in Greene 2000, p. 23) if rank is less than $\min(g, K)$.

The problem of *testing* for reduced rank in regression matrices is relatively new to the econometrics literature. Reduced-rank regression (RRR) in econometrics appeared as early as the 1970's, in structural equation modelling, in the case where the model's reduced form consists of incomplete simultaneous equations (for review see Kleinbergen 1999). Here the reduced-rank restriction is not usually regarded as a hypothesis to be tested per se, but is a consequence of market equilibrium or some other basic economic assumption. Yet tests for reduced rank are a valid and potentially important form of specification analysis (also 'path' analysis, see Kline 1998) for such models.

The vector autoregression model (VAR) has recently seen many applications of RRR. Here, the reduced rank restriction is frequently interpreted in terms of common trends (and co-integration, error correction), common cycles and related features of time series data (see

Johansen 1988, 1991, Stock and Watson 1988, Vahid and Engle 1993). Building on the seminal work of Anderson (1951) on RRR likelihood ratio (LR) tests, for Gaussian VAR models several authors have developed tests of the RRR restriction, including Johansen (1988,1991), Ahn and Reinsel (1988,1990) and Ahn (1997) (and see Kleinbergen 1999 for RRR score tests in VAR's with error heteroskedasticity).

Beyond the VAR model, economic theories such as the consumption-based capital asset pricing model (CAPM), the conditional CAPM, and the arbitrage pricing theory (see Cochrane 2001) have been shown to imply reduced rank in dynamic regressions of asset returns, as studied by Campbell (1987), Ferson and Foerster (1994), Bekker, Dobbelstein and Wansbeek (1996), and Costa, Gardini and Paruolo (1997). For these cases, Anderson's LR test can be applied (as in Bekker et al. 1996 and Costa et al. 1997), but if regression errors exhibit heteroskedasticity then other techniques are called for, such as heteroskedasticity-robust score tests (as in Campbell 1987 and Ferson and Foerster 1994).

Many other applications of RRR remain to be explored, and in the context of cross-section or panel data one can use reduced rank to specify, estimate and test forms of limited heterogeneity among groups in the population. This is true whether regressors are numerically identical across groups (as is assumed in most RRR testing), or not. We give two numerical examples, one with (cross section) regressions of men's and women's income on education levels, and one with (contemporaneous time series) regressions of large-firm and small-firm asset returns on economic factors/state variables. In the first example, the (pseudo)panel data is unbalanced, contains differing regressor values for men and women, and exhibits residual heteroskedasticity, creating difficulties both in practical implementation and theoretical justification of Anderson's LR test. In the second example, regressor values are the same across firms, but residual heteroskedasticity and serial correlation arise, once again calling for alternative tests.

Because the testing of RRR is relatively novel to econometrics, there have only recently been attempts to develop general tests, applicable to a wide variety of cases. Gill and Lewbel (1992) develop a Wald test of the rank of asymptotically normal matrices, and their approach can be considered very general for testing matrix rank, except in regard to unit root (and related sorts of) time series regression, where asymptotic normality can fail. Since the majority of multivariate regression analyzes in econometrics invoke assumptions sufficient for those of Gill and Lewbel (1992), for succinctness we go with this framework, although simulation results (discussed later) suggest that the proposed tests (with suitably modified critical values) are also useful for the co-integration case.

With an increasingly flexible testing framework, there remains a difficulty with RRR tests in econometrics, namely that the tests are all much more complicated than are popular tests of linear restrictions on regression models. By far the most popular approach to testing linear restrictions in regression is the F test method, in which one obtains unrestricted estimates of regression coefficients, and judges the proximity of the estimates to hypothesized parameter values. This approach is both mathematically precise and illustrative of a particularly appealing decision process for evaluating hypotheses.

As a simple approach to reduced rank testing, the econometrician can obtain unrestricted parameter estimates and then visually or graphically compare the coefficient matrix rows (or columns) to assess their redundancy, with 'comparisons' guided by supplementary statistics such as standard errors. To assess redundancy one can select a reference set of matrix rows and, for each remaining row, find the proximity of that row to (linear combinations of members within) the reference set. This heuristic method is viable so long as an appropriate reference set or basis is found, and *in that case* it is much simpler conceptually than the Wald test of Gill and Lewbel (1992). As a formal analog to this heuristic method, we utilize the distance between an unrestricted matrix estimate and an estimate constrained by reduced rank, in a suitable metric. The resulting test statistic is rather unconventional, being neither a score, likelihood ratio, or Wald statistic, and is instead an instance of the

Szroeter (1983) class of 'generalized Wald' statistics. This Szroeter-type test can also be viewed as a Hausman (1978) test, and Gouriéroux, Monfort and Renault (1993) study a Szroeter-type test in a related but different case where the model setup is that of Anderson (1951) (except error normality is not assumed), more restrictive in terms of error properties and regressor design, but less restrictive in terms of normalizations (bases, reference sets) used.

Two distance minimization problems emerge, based on equivalent formulations of the null hypothesis. In each case, the RRR restriction is of 'mixed-form' (Gouriéroux and Monfort 1989) in the parameters of interest and some auxiliary parameters. In the first case the mixed-form is 'explicit' and therefore obvious, while in the second, equivalent, case it is 'implicit'. For each method we derive estimators and tests. The tests are asymptotically chi square under the null hypothesis, and while this result can in principle be derived as an application of Szroeter's (1983) general theory (see also Gouriéroux and Monfort 1989), in our case we are able to take a much simpler expositional approach. We show further that the proposed RRR estimators are asymptotically normal, and we develop consistent estimators of their standard errors and variance-covariance matrices.

We use simulation to study the finite-sample performance of the proposed Szroeter-type RRR tests. We describe the fidelity of the exact test distribution to the inexact but asymptotically valid chi square approximation, for various sample sizes, the number g of groups, the number K of regressors, the rank of β, design of x, choices of estimation method for the unrestricted regression model, and choices of variance-covariance estimator for the unrestricted regression coefficient estimates. Smaller sample sizes, larger g and larger K typically lessen fidelity to the chi square distribution, and the simulations identify particular scenarios where fidelity is good, and where it breaks down.

The remainder of the paper is organized as follows. Section 2 describes the model, Section 3 gives two examples, Section 4 develops tests, and Section 5 studies the tests in simulation.

Section 6 illustrates the methods using the examples from Section 3, and Section 7 concludes.

2. Model

We consider the multivariate linear regression model:

$$y_{ij} = \beta_i\, x_{ij} + \psi_i\, z_{ij} + \varepsilon_{ij}, \quad i = 1, ..., g,\ j = 1, ..., n_i, \tag{1}$$

with $g \geq 2$ groups and n_i observations in the i-th group, $i = 1, ...g$. The variables $y_{i1}, ..., y_{in_i}$ are the observations for the i-th dependent variable, x_{ij} is the observed value for a $K \times 1$ vector, β_i is a $1 \times K$ vector, z_{ij} is the observed value for an $L \times 1$ vector, ψ_i is a $1 \times L$ vector, and ε_{ij} is a random error for which $E[\varepsilon_{ij}|x_{ij}, z_{ij}] = 0$ almost surely, for all i and j.

Let the $g \times K$ matrix β be the intended target of reduced rank restrictions, and let ψ be unrestricted. If the rank of β is $\min(g, K)$ then it is said to have full rank, otherwise it has reduced rank ($= q$, say). In the case $g \leq K$ it is helpful to partition β as:

$$\beta = \begin{bmatrix} \beta_a \\ \beta_b \end{bmatrix}, \tag{2}$$

where β_a is the sub-matrix of β consisting of the first q rows, and β_b is the sub-matrix consisting of the last $g - q$ rows. Here, reduced rank is more specifically reduced *row* rank, e.g. the rows of β are linearly dependent. If, for the true value β^* of β, the rows are spanned by first q rows then we have:

$$\beta_b^* = A\beta_a^*, \tag{3}$$

for some $(g - q) \times q$ matrix A. If β^* has row rank q but the rows of β^* are not spanned by it's first q rows, there is a reassignment of group numbers $i = 1, ..., g$ such that the first q rows will span the rest, in which case (3) holds. The actual assignment of group numbers is then a potentially important issue, particularly if $q > 1$. In the case $g \geq K$ of reduced *column* rank, the transpose matrix $(\beta^*)'$ has reduced row rank, and the proposed tests can then applied to $(\beta^*)'$. Hence, we will assume without loss of generality that $g \leq K$.

A second useful way to express the reduced-rank restriction (3) is:

$$\beta^* = \begin{bmatrix} I_q \\ \gamma \end{bmatrix} \lambda, \tag{4}$$

for some full-rank $(g - q) \times q$ matrix γ, and some full-rank $q \times K$ matrix λ. In particular, $\gamma = A$ and $\lambda = \beta_a^*$. Using the terminology of Gouriéroux and Monfort (1989), both (3) and (4) are *mixed form* restrictions on β and auxiliary parameters (A, γ, λ), and the form (3) is in *explicit* mixed form, while (4) is in *implicit* mixed form.

The linear model (1), subject to (3) (equivalently (4)), defines a broad class of reduced-rank regression (RRR) models, in some respects more general than has been previously considered in the econometrics literature, and somewhat more general than we will ultimately consider in the present work. We will additionally assume the availability of an estimator $\hat{\beta}$ of β^* and, with $n = (n_1, ..., n_g)$, an estimator $\hat{\Omega}_n$ of the variance-covariance matrix Ω_n of $vec^*(\hat{\beta})$, where $vec^*(\hat{\beta}) = vec(\hat{\beta}')$ is the $gK \times 1$ vector $(\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_g)'$, with $\hat{\beta}_1, ..., \hat{\beta}_g$ the rows of $\hat{\beta}$. We introduce the vec^* notation because $vec^*(\beta)$ is naturally partitioned under (3). We further impose the following assumption, with $\theta = vec^*(\beta)$ and notation $n \to \infty$ meaning $n_i \to \infty$ for $i = 1, ..., g$:

Assumption 1: Each of the following holds:

(i). $\hat{\beta}$ *is a full rank matrix, almost surely,*

(ii). Ω_n *and* Ω_n^{-1} *exist and are finite, for each* n*,*

(iii). $\Omega_n \to 0$ *as* $n \to \infty$*,*

(iv). $\Omega_n^{-1/2}(\hat{\theta} - \theta^*) \to N(0, I_{gK})$ *as* $n \to \infty$*,*

(v). $\Omega_n \hat{\Omega}_n^{-1} \to I_{gK}$ *as* $n \to \infty$*.*

For the estimator $\hat{\beta}$ we can, for example, use ordinary least squares (OLS, equation by equation) or seemingly unrelated regressions (SUR), and for the covariance matrix estimator $\hat{\Omega}$ we have a variety of choices common in economics applications, including the OLS and SUR covariance estimators, and several non-parametric forms including the White nonparametric heteroskedasticity-consistent covariance estimator (denoted White), the Newey-West nonparametric heteroskedasticity and autocorrelation consistent (HAC) covariance estimator, based on the Bartlett kernel and the data-dependent Newey and West (1994) bandwidth, with and without pre-whitening (denoted NW and NW-P, respectively). Among HAC methods we will also examine the quadratic spectral kernel with the Andrews (1991) data-dependent bandwidth (without pre-whitening, denoted A), and the Andrews-Monahan (1992) method (denoted AM) with pre-whitening.

Assumption 1 includes many cases which lie outside the *classical* reduced-rank regression framework (e.g. Anderson 1951 and Gouriéroux et al. 1993) where the data panel is balanced (e.g. $n_1 = n_2 = \cdots = n_p = n^*$ for some n^*), the x values are the same across groups, e.g.

$$x_{ij} = x_j^*, \quad i = 1, ..., g, \;\; j = 1, ..., n^*, \tag{5}$$

for some iid $(x_j^*, j = 1, ..., n^*)$, independent of regression errors which themselves are iid. We want to carry out RRR tests, in as simple a way as possible, for cases which include unbalanced panels, residual heteroskedasticity and autocorrelation, and group-heterogeneous regressor values. Regarding (5), in this setting the $q \times 1$ vector $u_j^* \equiv \lambda\, x_j^*$ is the j-th observation on a set of q latent factors, with the simplification $E[y_{ij}|x_{ij}, z_{ij}] = E[y_{ij}|u_j^*, z_{ij}]$. In more general terms, to interpret a multivariate regression with reduced-rank β it is desirable that the observations x_{ij} are for variables having the same relative *meaning* for each group i, yet the numerical *values* of x_{ij} may differ across groups, hence violating (5).

3. Examples

For a rather novel example of RRR, with an unbalanced (pseudo) panel, regressor values heterogeneous across groups, and conditional heteroskedasticity, consider:

Example 1 (Income, Gender and Education): Let there be $g = 2$ groups of U.S. workers, the first group male, the second female. For a random cross-section of workers, with n_1 males and n_2 females, let y_{ij} be the income of worker j in the i-th gender group. Historically, men have tended to earn higher incomes (on average) than women, and this remains true when averages are computed separately for different education levels, ages, etc. (see for example Borjas 2000, Figure 7-7). To further describe this gender gap, let $x_{ij} = (D_{ij}^{nh}, D_{ij}^{h}, D_{ij}^{b}, D_{ij}^{b+})$, where D_{ij}^{n} is a dummy variable indicating the worker has no high school diploma or higher degree, D_{ij}^{h} is a dummy variable indicating a high school (but no higher) diploma, D_{ij}^{b} is a dummy variable for a bachelor's degree (with no further education), and D_{ij}^{b+} is a dummy variable for workers with some post-bachelor's education. If $L = 0$ (and hence z is void) then β_1 and β_2 are vectors of mean male and female incomes, by education level. A relevant class of restrictions is:

$$\beta_2 = c\,\beta_1, \tag{6}$$

for some unobserved scalar c. With gender 'equality' signified by $c = 1$, if instead $c < 1$ then there is gender 'inequality' favoring males, and this equality is realized in a proportionately higher mean income for males across all education levels. If $\beta_2 < \beta_1$ (e.g. $\beta_{2j} < \beta_{1j}$ for $j = 1, ..., 4$) but (6) fails then the interpretation is income inequality with higher mean income for males and with disproportionately higher excess income for males (relative to females) at some education levels, relative to some other education levels. If we modify x_{ij} to contain an intercept term, e.g. $x_{ij} = (1, D_{ij}^{h}, D_{ij}^{b}, D_{ij}^{b+})$, then the resulting β continues to satisfy (6). The regression errors ε_{ij} *could* in principle be conditionally homoskedastic, but we find (in Section 6) instead evidence of conditional heteroskedasticity.

For an example of RRR with residual heteroskedasticity and autocorrelation, consider:

Example 2 (Asset Returns, Firm Size and Factor Pricing): There are $n = 2$ groups, the first consisting of firms having small capitalization, the second consisting of large-capitalization firms. The firms' values are observed at times $j = t = 1, ..., T = n^*$, with y_{ij} the excess return for the i-th firm type during the period $[j-1, j]$. Historically, small firms have tended to earn higher excess returns (on average) than have large firms, and to further describe this gap let $x_{ij} = x_j^*$ consist of seven macroeconomic variables: A value-weighted market return, a default premium, a term premium, and growth rates of consumption, industrial production, consumer prices and money supply. With this specification, we can interpret the model (1) as a *linear factor pricing model* (Cochrane 2001, p. 80), which encompasses the Sharpe-Lintner CAPM (factor = market return), the Consumption-Based CAPM (factor = consumption growth), plus a variety of other possibilities (similar to Chen, Roll and Ross 1986, as discussed in Section 6). The coefficient restriction (6), with $c > 1$, defines a class of relatively simple, single-latent-factor explanations for the higher mean return of small firms, with proportionately higher sensitivity to each priced (e.g. $\beta_{ij} \neq 0$ for $i = 1, 2$) economic factor. If (6) fails then the situation is more complex: At least two economic factors must be priced, and small firms must have disproportionately greater excess sensitivity (relative the large firms) to one economic factor than to some others. While the errors ε_{it} *could* in principle be serially uncorrelated and conditionally homoskedastic, empirical evidence (discussed in Section 6) suggests otherwise.

4. Tests and Estimators

To obtain relatively simple tests for reduced row rank in coefficient matrix β^*, subject to the normalization (3), it is helpful to regard the problem as one of inquiring whether the last $g-q$ rows of β are redundant, meaning that each can be obtained as linear combinations of the first q rows. This mental exercise is particularly simple when $q=1$, and the case $q=2$ is only somewhat harder, etc. To proceed, a simple approach is to inspect the unconstrained estimate $\hat{\beta}$, which from our earlier assumptions is asymptotically normal, and gauge the proximity of each of the last $g-q$ rows of $\hat{\beta}$ to the span (linear combinations) of the first q rows. Alternatively, one can compare *all* rows to proxies obtained from a reduced-rank estimate, using (4).

To formalize the heuristic test procedure, let M_{gK} be the set of $g \times K$ matrices, and define a distance function on $M_{gK} \times M_{gK}$:

$$d(a,b;Q) \;\;=\;\; \Big[\,(vec^*\,a - vec^*\,b)'\,Q\,(vec^*\,a - vec^*\,b)\,\Big]^{1/2}, \tag{7}$$

for each a and b in M_{gK}, and some symmetric positive definite matrix Q, in which case $d(a,b;Q)$ is a metric on $M_{gK} \times M_{gK}$. We will define reduced-rank restricted estimators $\tilde{\beta}$ by minimizing $d^2(\hat{\beta},\beta,Q)$ over suitable sets of β. For testing reduced rank, we will define test statistics of the form:

$$W = d^2(\hat{\beta},\tilde{\beta};Q), \tag{8}$$

which are a subclass of Szroeter's statistics "W" (Szroeter 1983, equation 4.16). The proposed tests will reject the hypothesis rank $(\beta) = q$ if and only if W exceeds the relevant critical value from the chi square distribution, with $(g-q)(K-q)$ degrees of freedom. The tests will be consistent against alternatives rank$(\beta) > q$, but not against alternatives where rank $< q$, hence the implicit null hypothesis includes the latter possibilities. Also, in the mixed-form rank $= q$ parameterizations (3) and (4), if in reality the rank $(\beta) < q$ then some parameters are not identified, and hence W takes on a nonstandard distribution (see Andrews and Ploberger 1994) in that case.

To proceed, in Section 3.1 and 3.2 we will repeatedly make use of some facts about the row-stacking vec^* operator. For any $a \times b$ matrix M_1 and $b \times c$ matrix M_2, we have:

$$vec^*(M_1 M_2) = (M_1 \otimes I_c)\, vec^*(M_2) = (I_a \otimes M_2')\, vec^*(M_1). \tag{9}$$

To obtain (9) we note that, for any matrix C, $vec^*(C) = vec(C')$, and we then apply standard rules for $vec(C)$ when $C = M_1 M_2$, as in Ruud (2000, p. 925).

4.1. Explicit Mixed-Form

Let S_q the set of $g \times K$ matrices β of the form:

$$\beta = \begin{bmatrix} \hat{\beta}_a \\ A\hat{\beta}_a \end{bmatrix}, \tag{10}$$

for some $(g-q) \times q$ matrix A. Then, provided that $\hat{\beta}_a$ has rank q (which occurs almost surely, by Assumption 1), each element of S_q has rank q, and we define an estimator $\tilde{\beta}$:

$$\tilde{\beta} = \text{argmin}_{\beta \in S_q}\, d^2(\hat{\beta}, \beta; Q). \tag{11}$$

To compute $\tilde{\beta}$, note first that $\tilde{\beta}_a = \hat{\beta}_a$. Also, for each β satisfying (10) we have:

$$d^2(\hat{\beta}, \beta; Q) = \left(vec^*\,\hat{\beta}_b - vec^*\,(A\hat{\beta}_a)\right)' Q_{bb} \left(vec^*\,\hat{\beta}_b - vec^*\,(A\hat{\beta}_a)\right), \tag{12}$$

where Q_{bb} is the lower-right $(g-q)\times(g-q)$ sub-matrix of Q. Using (9), we have $vec^*(A\hat{\beta}_a) = (I_{g-q} \otimes \hat{\beta}_a')\, vec^*(A)$, in which case, for each β satisfying (10), we can then regard $d^2(\beta, \hat{\beta}, Q)$ as a function $g(z)$ of the $q(g-q) \times 1$ vector $z = vec^*(A)$. The problem is then to minimize $g(z)$ over the set of $z \in R^{q(g-q)}$. Since $g(z)$ is differentiable, the necessary first-order conditions for a minimum are as follows, with $dg(z)/dz = (\partial g(z)/\partial z_1, ..., \partial g(z)/\partial z_{q(g-q)})$:

$$0 = \frac{dg(z)}{dz} = -2\left(vec^*\,\hat{\beta}_b - (I_{g-q} \otimes \hat{\beta}'_a)z\right)' Q_{bb}(I_{g-q} \otimes \hat{\beta}'_a), \tag{13}$$

in which case:

$$\tilde{z} = \left[(I_{g-q} \otimes \hat{\beta}'_a)' Q_{bb}(I_{g-q} \otimes \hat{\beta}'_a)\right]^{-1} (I_{g-q} \otimes \hat{\beta}'_a)'\, Q_{bb}\, vec^*(\hat{\beta}_b). \tag{14}$$

With $\tilde{A}$ the estimator of A defined by $\tilde{z} = vec^*(\tilde{A})$, we obtain $\tilde{\beta}$ by replacing A with $\tilde{A}$ in (10). Moreover, we have:

$$\frac{d^2 g(z)}{dz^2} = 2(I_{g-q} \otimes \hat{\beta}'_a)' Q_{bb}(I_{g-q} \otimes \hat{\beta}'_a), \tag{15}$$

and since $\hat{\beta}_a$ is almost surely of full rank (and Q_{bb} is assumed invertible), $\frac{d^2 g(z)}{dz^2}$ is positive semi-definite, a sufficient condition for $\tilde{z}$ to be a minimum of $g(z)$, hence $\tilde{\beta}$ solves (11).

We now turn to the choice of the matrix Q_{bb}. By Assumption 1, $\hat{\beta}$ is normal asymptotically. Applying (9), we have $vec^*(A\hat{\beta}_a) = (A \otimes I_K) vec^*(\hat{\beta}_a)$, giving the asymptotic (large n) approximation:

$$vec^*(\hat{\beta}_b) = B\, vec^*(\hat{\beta}_a) + \zeta, \tag{16}$$

where $B = A \otimes I_K$, and ζ is distributed $N(0, V_\zeta)$. Furthermore,

$$V_\zeta \;=\; E\,(\,vec^*(\hat{\beta}_b) - B\, vec^*(\hat{\beta}_a)\,)(\,vec^*(\hat{\beta}_b) - B\, vec^*(\hat{\beta}_a)\,)', \tag{17}$$

and hence:

$$V_\zeta \;=\; E\,vec^*(\hat\beta_b)[vec^*(\hat\beta_b)]' \;+\; B\,E\;vec^*(\hat\beta_a)[vec^*(\hat\beta_a)]'\,B'$$
$$-\;E\,vec^*(\hat\beta_b)[\,vec^*(\hat\beta_a)]'B' \;-\; B\,E\;vec^*(\hat\beta_a)[vec^*(\hat\beta_b)]'.$$

For succinctness, we can then express V_ζ as:

$$V_\zeta \;=\; C\,\Omega_n C', \tag{18}$$

with C the $((g-q)K) \times gK$ matrix:

$$C = \left[\; -A \otimes I_K\;,\quad I_{(g-q)K}\;\right]. \tag{19}$$

If the matrix V_ζ was observable, we could set $Q_{bb} = V_\zeta^{-1}$, in which case the minimization problem (11) would take the form of maximizing a(n) (approximate) normal likelihood for the density of $\hat\beta_b$, conditional on $\hat\beta_a$. But V_ζ is unobserved, and we set $Q_{bb} = \tilde V_\zeta^{-1}$, with $\tilde V_\zeta$ obtained by replacing A with a consistent estimate $A^\dagger$ in the definition of V_ζ, and replacing Ω_n with $\hat\Omega_n$. In particular, let $A^\dagger$ solve (11) with Q_{bb} equal to the lower-right $(g-q)K \times (g-q)K$ sub-matrix of $\hat\Omega_n^{-1}$.

Some important properties of the proposed estimators are as follows (see Appendix 1 for proofs):

Theorem 1: If Assumption 1 holds and β^* has rank q then $V_{\tilde z}^{-1/2}(\tilde z - z) \to N(0, I_{(g-q)q})$ in distribution, where $V_{\tilde z} = \left((I_{g-q} \otimes (\beta_a^*)')'V_\zeta^{-1}(I_{g-q} \otimes (\beta_a^*)')\right)^{-1}$.

Theorem 2: If Assumption 1 holds and β^* has rank q then $V_\zeta^{-1/2}(vec^*(\hat\beta_b) - vec^*(\tilde\beta_b)) \to N(0, I_{(g-q)K})$ in distribution.

We then have a test statistic:

$$W = (vec^*(\hat{\beta}_b) - vec^*(\tilde{\beta}_b))' \tilde{V}_\zeta^{-1} (vec^*(\hat{\beta}_b) - vec^*(\tilde{\beta}_b)), \tag{20}$$

of the form (8) with:

$$Q = \begin{bmatrix} 0 & 0 \\ 0 & \tilde{V}_\zeta^{-1} \end{bmatrix}.$$

Corollary 1: If Assumption 1 holds and β^* has rank $= q$ then, as $n \to \infty$, the statistic W defined by (20) converges in distribution to chi square, with $(g-q)(K-q)$ degrees of freedom.

4.2 Implicit Mixed-Form

Let U_q be the set of $g \times K$ matrices β of the form (4), for some $g \times q$ matrix γ and $q \times K$ matrix λ. We can then define a reduced-rank estimator:

$$\ddot{\beta} = \text{argmin}_{\beta \in U_q} d^2(\hat{\beta}, \beta, Q). \tag{21}$$

Under (4) we have $\beta = D\lambda$, with:

$$D = \begin{bmatrix} I_q \\ \gamma \end{bmatrix}.$$

For such β we can regard $d^2(\hat{\beta}, \beta, Q)$ as a function $h(z)$ with $z = ((vec^*(\gamma))', (vec^*(\lambda))')'$. Applying (9) we have $vec^*(D\lambda) = (D \otimes I_K)\, vec^*(\lambda) = (I_g \otimes \lambda')\, vec^*(D)$, necessary first-order conditions for a minimum of h are as follows, with $\partial h(z)/\partial z = (\partial h(z)/\partial z_1, ..., \partial h(z)/\partial z_{gq+Kq-q^2})$.

$$0 = \frac{\partial h(z)}{\partial vec^*(\gamma)} = -2(vec^*(\hat{\beta}) - (I_g \otimes \lambda')\, vec^*(D))'\, Q\, (I_g \otimes \lambda')E, \tag{22}$$

$$0 = \frac{\partial h(z)}{\partial vec^*(\lambda)} = -2(vec^*(\hat{\beta}) - (D \otimes I_K)\, vec^*(\lambda))'\, Q\, (D \otimes I_K). \tag{23}$$

where:

$$E = \begin{bmatrix} 0_{q^2,(g-q)q} \\ I_{(g-q)q} \end{bmatrix}.$$

Partial solutions for $vec^*(\gamma)$ and $vec^*(D)$ are then:

$$vec^*(D) = \left[((I_g \otimes \lambda')E)'\, Q\, (I_g \otimes \lambda')\right]^{-1} ((I_g \otimes \lambda')E)'\, Q\, vec^*(\hat{\beta}), \tag{24}$$

$$vec^*(\lambda) = \left[(D \otimes I_K)'\, Q\, (D \otimes I_K)\right]^{-1} (D \otimes I_K)\, Q\, vec^*(\hat{\beta}). \tag{25}$$

Second derivatives of h include:

$$\frac{\partial^2 h(z)}{\partial\, vec^*(\gamma) \partial\, vec^*(\gamma)'} = 2((I_g \otimes \lambda')\, E)'\, Q\, (I_g \otimes \lambda')E, \tag{26}$$

$$\frac{\partial^2 h(z)}{\partial\, vec^*(\lambda) \partial\, vec^*(\lambda)'} = 2(D \otimes I_K)'\, Q\, (D \otimes I_K). \tag{27}$$

We now turn to the choice of the matrix Q. Since, by Assumption 1, $\hat{\beta}$ is normal asymptotically, we have the asymptotic (large n) approximation:

$$vec^*(\hat{\beta}) = vec^*(D\lambda) + \eta, \tag{28}$$

with η distributed $N(0, \Omega_n)$. If Ω_n was observable then the choice $Q = \Omega_n^{-1}$ would cause the problem (21) to be that of maximizing an approximate normal likelihood for $\hat{\beta}$. We do not observe Ω_n, and propose instead $Q = \hat{\Omega}_n^{-1}$.

Estimator distributions are as follows:

Theorem 3: If Assumption 1 holds and β^* has rank q and $Q = \hat{\Omega}_n^{-1}$ then each of the following is true of matrices $\ddot{\gamma}$ and $\ddot{\lambda}$ solving (21):

(i) $V_\gamma^{-1/2}(vec^*(\ddot{\gamma}) - vec^*(\gamma^*))$ *is distributed asymptotically as* $N(0, I_{(g-q)K})$,

where $V_\gamma = ((\,I_{g-q} \otimes \lambda')' \,\hat{\Omega}_n^{-1}\, I_{g-q} \otimes \lambda'\,)^{-1}$.

(ii) $V_\lambda^{-1/2}(vec^*(\ddot{\lambda}) - vec^*(\lambda^*))$ *is distributed asymptotically as* $N(0, I_{(qK)})$,

where $V_\lambda = ((D \otimes I_K)' \,\hat{\Omega}_n^{-1}\, (D \otimes I_K))^{-1}$.

We can compare, via Theorems 1 and 3, the implicit mixed-form estimator $\ddot{\beta}$ to the explicit mixed-form estimator $\tilde{\beta}$. With $\gamma = A$ and $\lambda = \beta_a^*$, the estimator $\ddot{\gamma} = \ddot{A}$ is more efficient than $\tilde{\gamma} = \tilde{A}$, and $\ddot{\lambda} = \ddot{\beta}_a$ is more efficient than $\tilde{\lambda} = \tilde{\beta}_a$.

Theorem 4: If Assumption 1 holds and β^* *has rank* q *then* $\hat{\Omega}_n^{-1/2}(vec^*(\hat{\beta}) - vec^*(\ddot{\beta})) \to N(0, I_{gK})$ *in distribution.*

We then have a test statistic:

$$W = (vec^*(\hat{\beta}) - vec^*(\ddot{\beta}))'\hat{\Omega}_n^{-1}(vec^*(\hat{\beta}) - vec^*(\ddot{\beta})). \tag{29}$$

Corollary 2: If Assumption 1 holds and β *has reduced rank then, as* $n \to \infty$*, the statistic* W *defined by (29) converges in distribution to chi square, with* $(g-q)(K-q)$ *degrees of freedom.*

To compute $\ddot{\beta}$ and the statistic W defined by (29), we use an iterative approach, generating a sequence $(\ddot{\gamma}^{(k)}, \ddot{\lambda}^{(k)}, k = 1, 2, ..., m)$, for some number m of iterations. We set $\ddot{\lambda}_1 = \hat{\beta}_a$, then we compute $\ddot{\gamma}_1$ from (24) by replacing λ with $\ddot{\lambda}_1$ in that formula. We then compute $\ddot{\lambda}_2$ from (25) by replacing γ with $\ddot{\gamma}_1$ in that formula. We continue the alternating use of (24) and (25) until iteration stops. In the examples we have study in Sections 5 and 6, we have found the choice $m = 10$ to be good in terms of algorithm (near)-convergence. In some extreme cases the algorithm may converge to a local (but non-global) optimum, and we have confirmed this in some cases where $\hat{\beta}$ is a diagonal matrix. Therefore, it may be helpful to try several starting points $(\ddot{\lambda}_1)$ for the algorithm.

5. Simulation

We simulate the finite-sample rejection rates of the proposed reduced-rank tests, for a variety of model specifications. We begin in Section 5.1 with a simple VAR model, then in Section 5.2 turn to a model where regressors are numerically different across groups/equations. Section 5.3 reports briefly on the case of co-integration.

5.1 VAR

For the VAR model we have::

$$x_{ij} = x_j^* = (1, y_{1,j-1}, ..., y_{g,j-1}, y_{1,j-2}, ..., y_{g,j-2}, ..., y_{1,j-p}, ..., y_{g,j-p})', \tag{30}$$

with p the number of lags, in which case $K = 1 + g\,p$. For simplicity, we specify the regression errors as independent and identically distributed normal $N(0, \Delta)$ over time, with Δ being either the identity matrix I_g or the $g \times g$ matrix Υ having 2's on the diagonal and the value 1 on the off-diagonal. Under the null hypothesis $q = 1$, we specify $\beta_{i2} = 0.25$ for each i and $\beta_{ij} = 0$ for $j \neq 2$. For example, the DGP in the simplest case $(g = 2, q = 1, K = 3)$ is:

$$\begin{aligned} y_{1j} = &\ 0.25 y_{1,j-1} + \varepsilon_{1j}, \\ y_{2j} = &\ 0.25 y_{1,j-1} + \epsilon_{2j}, \end{aligned} \tag{31}$$

for $j = p + 1, ..., n$. The unrestricted model to be estimated is then:

$$\begin{aligned} y_{1j} = &\ \beta_{11} + \beta_{12} y_{1,j-1} + \beta_{13} y_{2,j-1} + \varepsilon_{1j}, \\ y_{2j} = &\ \beta_{21} + \beta_{22} y_{1,j-1} + \beta_{23} y_{2,j-1} + \varepsilon_{2j}, \end{aligned} \tag{32}$$

for $j = p+1, ..., n$. Under the null hypothesis $q = 2$ (and $g = 3$), we specify $\beta_{12} = 0.25, \beta_{22} = 0.25, \beta_{33} = 0.25$, with all remaining β elements equal to 0. Under the alternative hypothesis $(q = g)$, we specify β such that each y series is an AR(1) series with drift term 0 and AR slope 0.25.

For each version of the calibrated model, we use a normal random number generator (in EViews 3.1, build 6/2000) to get a simulated realization of the y_j-process, $j = p + 1, \ldots, n$. For each simulation round we compute the proposed reduced-rank tests, and using the asymptotically valid (chi square) critical values, we determine whether each test rejects or fails to reject, at the 5 percent significance level. We repeat the simulation 500 times, and record the empirical frequency of rejection.

We first report on the proposed tests calculated using OLS and SUR multivariate regression estimators and standard errors. Tables 1 and 2 report the rejection rates for the explicit and implicit mixed-forms of the test, under the null and alternative hypotheses, for errors following either the standard normal distribution or the normal distribution with non-zero cross-covariances, as specified above. We use two sample sizes, 150 and 300, respectively. For the explicit mixed-form, under the null hypothesis empirical rejection rates sometimes depart noticeably from the nominal 5% level, at either sample size, but are generally closer to 5% for the larger sample (and even closer to 5% for samples of size 500, as we have verified but omit for brevity). For the implicit mixed-form, finite-sample distortions appear more mild, but also display lower rejection frequency under the alternative hypothesis.

The second batch of our simulation exercise attempts to address two issues. First, how do the proposed Szroeter-type tests compare to score tests? To investigate, we use score tests based on the method of moments, via the simultaneous-iteration procedure described in Hansen, Heaton and Yaron (1996), where parameter and covariance matrix estimates are updated simultaneously at each iteration stage. The second issue is the finite sample performance for RRR tests when heteroskedasticity-consistent and HAC methods are used to calculate the regression estimator variance-covariance matrices. Table 3 reports rejection rates. For brevity we report only for the sample size $n = 150$, for the implicit mixed-form Szroeter test and the score test, computed via the White, NW, NW-P, A an AM covariance estimators. For the explicit mixed-form Szroeter tests, results are similar to the implicit mixed-form, but with somewhat higher rejection rates under the null and alternative hypotheses. Simulations indicate reasonably good behavior (little distortion) under the null hypothesis for the simplest ($g = 2, K = 3, q = 1$) model. As the complexity of the model increases (g, K, q increase), we observe greater distortions, with the Szroeter test typically over-rejecting and the score test under-rejecting the null. Distortions are to be expected for at least two reasons, first because the reduced-rank restriction on *beta* is non-linear and hence the asymptotic chi square distribution is not exact, and second because we use heteroskedasticity-consistent or HAC methods which themselves are known to introduce distortions into the test distributions. As we can see from Table 3, the choice of the covariance estimator implies differences in the small sample performance of the tests, though they are often minor. Besides the aforementioned covariance estimators, we also examined the simple pre-whitening method studied by den Haan and Levin (1996, 1997), with parametric, VAR adjustment for autocorrelation. Since the results seem to parallel our results when AM method is used, we omit them for brevity.

5.2 Explanatory variables with differing values across

In our second simulation setup, we study the behavior of RRR tests in an environment where the explanatory variables have different numerical values across groups/equations. We set $x_{ij1} = 1$ and, for $k = 2, \ldots, 4$, $x_{ijk} = 1$ if the uniform random number generator gives a value between $0.25(k-1)$ and $0.25k$ and $x_{ijk} = 0$ otherwise. The dependent variables are constructed under the null hypothesis as follows:

$$\begin{aligned} y_{1j} = &\quad 1 + 2x_{1j2} + 3x_{1j3} + 4x_{1j4} + \varepsilon_{1j}, \\ y_{2j} = &\quad 0.8(1 + 2x_{2j2} + 3x_{2j3} + 4x_{2j4}) + \varepsilon_{2j}, \end{aligned} \tag{33}$$

for $j = 1, \ldots, 150$ with ε's iid standard normal. We repeat the experiment 500 times and report the empirical size in Panel A of Table 4. Results indicate that the finite-sample size of the explicit mixed-form Szroeter test is essentially identical to the theoretical size, while that of the implicit mixed-form test tends to over-reject, and the score test tends to under-reject.

The full rank alternative hypothesis is given by:

$$\begin{aligned} y_{1j} = &\ \ 1 + 2x_{1j2} + 3x_{1j3} + 4x_{1j4} + \varepsilon_{1j}, \\ y_{2j} = &\ \ 0.9(1 + 2x_{2j2}) + 0.7(3x_{2j3} + 4x_{2j4}) + \varepsilon_{2j}, \end{aligned} \tag{34}$$

again for $j = 1, \ldots, 150$ and with ε's iid standard normal. From Panel A of Table 4, Both Szroeter tests have higher rejection rates as compared to the score tests, with the explicit form dominating the implicit form even though the empirical size of the former is lower.

5.3 Co-integration

In our last simulation batch, we investigate the performance of the RRR tests in the co-integration case. The presence of unit roots and co-integration in the data changes distributional properties (Assumption 1) of estimators and resulting tests. We assume that the DGP under the null hypothesis is:

$$\begin{aligned} y_{1j} = &\ \ 1 + y_{1,j-1} + \varepsilon_1, \\ y_{2j} = &\ \ y_{1j} + \varepsilon_2, \end{aligned} \tag{35}$$

for $j = 1, ..., 150$ and ε's iid standard normal. The unrestricted system to be estimated is:

$$\begin{aligned} y_{1j} = &\ \ \beta_{11} y_{1,j-1} + \beta_{12} y_{2,j-1} + \psi_1 + \varepsilon_1, \\ y_{2j} = &\ \ \beta_{21} y_{1,j-1} + \beta_{22} y_{2,j-1} + \psi_2 + \varepsilon_2, \end{aligned} \tag{36}$$

for $j = 1, ..., 150$. Under the null hypothesis, we have

$$\beta = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

and it has a rank = 1. In the restricted system, we used the following normalization:

$$\begin{aligned} y_{1j} = &\ \ \beta_{11} y_{1,j-1} + \beta_{12} y_{2,j-1} + \psi_1 + \varepsilon_1, \\ y_{2j} = &\ \ \gamma(\beta_{11} y_{1,j-1} + \beta_{12} y_{2,j-1}) + \psi_2 + \varepsilon_2, \end{aligned} \tag{37}$$

for $j = 1, ..., 150$. For critical values we do not use the chi-square distribution but instead the distribution for λ_{trace} based on the Johansen method. In our case, we use the $\lambda_{\text{trace}} = 3.962$, as reported in Enders (1995, Table B, p. 420) for the 95% quantile for the case with trend drift. However, the asymptotic chi-square distribution for one degree of freedom is 3.84 and its use instead of 3.962 does not alter the distributional pattern in any significant manner. The results (with c.v. = 3.962) are shown in Panel B of Table 4. Both forms of the Szroeter test give similar results and over-reject with respect to the score test. For the rejection rates under the alternative (see Panel B, Table 4) we employ the model:

$$\begin{aligned} y_{1j} &= \ 1 + y_{1,j-1} + \varepsilon_1, \\ y_{2j} &= \ 1 + 0.8y_{1,j-1} + 0.2y_{2,j-1} + \varepsilon_2, \end{aligned} \tag{38}$$

for $j = 1, ..., 150$ and standard normal ε's. Rejection rates are consistently higher for the Szroeter tests than for the score test.

6. Examples, Cont'd

Example 1 - Income, Gender and Education

We use data from Integrated Public Use Micro-data Samples database, available at http:// www.ipums.umn.edu (see Ruggles, Sobek et al. 1997 for a description of the dataset). We extracted a random sample from 1990, with variables sex, total income (INCTOT) and education (EDUC99). INCTOT is an individual's total pre-tax personal income or loss from all sources for the previous calendar year. EDUC99 variable has 17 categories, which we reduce to only four: no diploma, a high school graduate, a college graduate, a post-bachelor's degree. Our choice of the education variable reflects evidence of the importance of the achieved degree over years of schooling (the so called 'sheepskin effect', see for instance Jaeger and Page 1996). We consider persons 16 years old and older and have $n_1 = 2105$ observations for males and $n_2 = 2289$ for females on both income and education. Summary statistics are reported in Panel A, table 5. The pattern of the gender pay-gap is similar to the one in Borjas (2000, Figure 7-7). For the sake of simplicity, we limit ourselves to the use of education dummies as the only explanatory variables. Other potential explanatory variables would include experience, age, quality of schooling, etc. For other gender-specific factors, see Becker (1993) and more recently, Blau and Kahn (2000).

We estimate the multivariate regression model (1) by ordinary least squares, equation by equation. The White heteroskedasticity test indicates presence of heteroskedasticity in both equations (the F statistics are 41.65 for males and 63.44 for females, with p-values = 0.000). The unrestricted coefficient estimates ($\hat{\beta}$), with White heteroskedasticity-robust standard errors are reported in Panel B, Table 5. To examine the reduced-rank hypothesis, we use only the information in Panel B, and for the coefficient variance-covariance estimate $\hat{\Omega}$ we let diagonal elements of $\hat{\Omega}$ equal the squared standard errors for the relevant coefficient, and we let the off-diagonal elements of $\hat{\Omega}$ equal 0 (consistent with our regressor design).

A casual inspection suggests that, in 1990, women's average income was about half that of men's, for each education level. The proposed RRR estimators and tests, reported in Panel B, agree with this conclusion (and we omit here a report on method-of-moments estimators and score tests here because the unbalanced sample creates some difficulties for these methods in standard programs like EViews). The multiplier A, from men's to women's income, is estimated near 0.5 by the explicit and implicit mixed-form methods, and both forms of the Szroeter test fail to reject the rank = 1 hypothesis for β^*. Only the implicit mixed-form method provides *revised* estimates of men's mean incomes (by education level), and from the reported standard errors these have somewhat less sampling error than the initial (OLS) estimates.

Example 2 - Asset Returns, Firm Size and Factor Pricing

For the small-firm and large-firm portfolio asset returns, we use NYSE Cap-Based Portfolio Indices provided by CRSP. We compute excess returns using the 30-Day Treasury Bill return, also from CRSP, for the period 1959:02 - 1999:12. We denote the excess returns by r_{SMALL} and r_{LARGE}. We use the return on the CRSP NYSE value-weighted index, in excess of the 30-Day Treasury Bill, as a measure of overall stock market performance (denoted r_{VW}). The default premium (r_{DEF}) is the difference between interest rates on low grade bonds (the Seasoned Baa Corporate Bond Yield from the St. Louis FED website, the original source is Moody's Investors Service) and long-term government securities (5-year Treasury Bonds, the St. Louis FED website). The term premium (r_{TERM}) is the difference between the one-period holding return on the 5-year Treasury Bond (CRSP) and the first lag of the return on a 30-Day Treasury Bill. For the growth rates of industrial production (g_{IP}) and real per capita consumption (g_{CONS}), the industrial production data are obtained from the Federal Reserve Board's website (Market Groups, series b50001, seasonally adjusted) and consumption data from the St. Louis Fed's website (series PCEND, non-durables, series PCES, services, POP, population, series CPIAUCSL, Consumer Price Index For All Urban Consumers, All Items 1982-84=100, all series seasonally adjusted). As in Chen, Roll, et al. (1986), we measure the effects of inflation using the unexpected inflation(π_{UI}), calculated as the difference between actual inflation rate (measured by the above consumer price index) and a one-period forecast of the inflation rate. The one-period forecast of the inflation rate is estimated using coefficients from a regression of the inflation rate, a constant, its lagged value, the lagged value of a Treasury Bill rate and a moving average term. As a measure of the money supply, we use the growth rate of the seasonally adjusted monetary base from the St. Louis Fed's website(series AMBSL, seasonally adjusted, denoted g_{MON}). Panel A of Table 6 summarizes the data series.

Panel B of Table 6 reports unrestricted coefficient estimates, and because there is evidence of conditional heteroskedasticity and autocorrelation (see Panel D), we use a suitably robust form (AM = Andrews-Monahan) of standard errors. Panel B suggests that several coefficient estimates ('sensitivities', for the default premium, term premium, consumption growth, money growth) differ substantially across firm groupings, with the exception of the market beta, which is close to 1 for both small and large firms. This is contrary to the idea that the large-firm coefficient vector β_b^* is a scalar multiple of the small-firm coefficient vector β_a^*. While we omit (for brevity) a report of covariances among coefficient estimates, these too will be used in our Szroeter-type tests for reduced rank in β^*.

Panel C reports the proposed Szroeter-type RRR tests and estimators, and also reports on RRR method of moments estimators and score test statistics (again using the AM specification). Each test rejects the rank = 1 hypothesis, at the 1 percent significance level. An interpretation is that, among the chosen economic factors, small and large firms are sensitive to several of these factors, and small-firm sensitivities are disproportionately high for some factors (default premium), relative to large firms, and disproportionately low for some other factors (term premium, consumption growth, money growth). This description of the link between firm size and asset returns is only illustrative (for more on this issue, see Schwert 1983 for discussion of early theories, and Fama and French 1992, 1993 provide recent empirical treatments), but it further demonstrates the utility of the RRR methods.

7. Discussion

In this paper we consider the problem of estimation and testing for reduced rank in multivariate linear regression models. We propose estimators and tests which formalize the intuitive approach of examining redundancy in an unconstrained estimate of the regression coefficient matrix. We verify the asymptotic normal and chi square distributions of the proposed statistics, and we conduct simulations to assess the finite-sample properties of the proposed tests. In simulation we find that rejection rates under the hypotheses of both full and reduced ranks are similar for both the explicit and implicit mixed-forms of the test, and both tests produced coherent results, provided that the sample was not too small.

The proposed methods are suited to the case where there is a known basis of regression coefficient rows (resp. columns) that span all the rows (columns). This is valid in the simplest case (rank = 1) of reduced rank regression, but otherwise calls for some justification. Of the two proposed methods, each based on minimum distance estimation of parameters subject to explicit and implicit mixed-form restrictions, the explicit form is simpler and is particularly tied to the specified basis for coefficient rows, whereas the implicit mixed-form can be easily generalized to allow any basis (by allowing all elements of D and λ in the representation $\beta = D\lambda$ to freely vary). Hence, a slightly modified version of our implicit mixed-form approach achieves robustness to choice of basis, and in simulation (omitted for brevity) is essentially identical to the original implicit approach in the cases studied in Section 5. However, the modified implicit version is also more complex conceptually.

Future work could further compare the power of the explicit and implicit mixed-form tests, using local power asymptotics. Since the implicit mixed-form estimators include relatively efficient estimators of *all* parameters (in β, under reduced rank), compared to the explicit approach, there may be a local power advantage to the implicit mixed-form tests. Also interesting would be a derivation of formulas for second-order asymptotic bias and variance for the proposed estimators, as well as theories modified for the case of unit roots and co-integration.

APPENDIX

Proof of Theorem 1: With z^* the true value of $z = vec^*(A)$ under the condition rank $(\beta^*) = q$, we can express $\tilde{z}$ as a differentiable function $\phi(vec^*(\hat{\beta}), hvec(\tilde{V}_\zeta))$ of $vec^*(\hat{\beta})$ and $hvec(\tilde{V}_\zeta)$. As $n \to \infty$, we can then apply a first-order Taylor series Approximation to obtain:

$$\tilde{z} - z^* = G\,(vec^*(\hat{\beta}) - vec^*(\beta^*)) + H\,(hvec(\tilde{V}_\zeta) - hvec(V_\zeta)) + o,$$

where G and H are matrices of partial derivatives of ϕ with respect to it's first and second set of arguments (evaluated at their population values), and where "o" indicates a negligible (second order) term. Moreover, using the formula (14) we deduce that $H = 0$, and hence $\tilde{z} - z^*$ is asymptotically equivalent to $G\,(vec^*(\hat{\beta}) - vec^*(\beta^*))$. From this, the asymptotic normality of $\hat{\beta}$ implies that $\tilde{z}$ is also asymptotically normal. The proposed variance-covariance formula $V_{\tilde{z}}$, given by $\left((I \otimes (\beta_a^*)')' V_\zeta^{-1} (I \otimes (\beta_a^*)')\right)^{-1}$ is valid (for the purpose of Theorem 1) due to the asymptotic normality of $\hat{\beta}$ and the fact that $\tilde{z}$ maximizes the normal likelihood (16) (conditional on the consistent estimator $\tilde{V}_\zeta$ of V_ζ).

Proof of Theorem 2: Follows from the asymptotic normality of $\hat{\beta}$ and the fact that $\tilde{z}$ maximizes the normal likelihood (16) (conditional on the consistent estimator $\tilde{V}_\zeta$ of V_ζ).

Proof of Theorem 3: For the vector $\ddot{v} = (vec^*(\ddot{\gamma})', vec^*(\ddot{\lambda})')'$ of parameter estimates, we can proceed analogous to the proof of Theorem 1 to express $\ddot{v}$ as a differentiable function, say $\kappa(\cdot)$, of $\hat{\beta}$ and $\hat{\Omega}$, and then obtain a first-order Taylor series Approximation which, upon simplification, yields $\ddot{v} - v^* = J(vec^*(\hat{\beta}) - vec^*(\beta^*))$, with J a matrix of partial derivatives of κ with respect to the elements of $vec^*(\hat{\beta})$. Consequently, since $\hat{\beta}$ is asymptotically normal, so is $\ddot{z}$. The proposed variance-covariance formulas for $\ddot{\gamma}$ and $\ddot{\lambda}$ are valid (for the purpose of Theorem 3) due to the asymptotic normality of $\hat{\beta}$ and the fact that $\ddot{v}$ maximizes the normal likelihood (28) (conditional on the consistent estimator $\hat{\Omega}_n$ of Ω_n).

Proof of Theorem 4: Follows from the asymptotic normality of $\hat{\beta}$ and the fact that $\ddot{v}$ (defined in proof of Theorem 3) maximizes the normal likelihood (28), conditional on the consistent estimator $\hat{\Omega}_n$ of Ω_n.

References

Ahn, S.K., 1997. Inference of Vector Autoregressive Models with Cointegration and Scalar Components. Journal of the American Statistical Association 92, 350-356.

Ahn, S.K., Reinsel, G.S., 1988. Nested Reduced-Rank Autoregressive Models for Multiple Time Series. Journal of the American Statistical Association 83, 849-856.

Ahn, S.K., Reinsel, G.C., 1990. Estimation of Partially Nonstationary Multivariate Autoregressive Models. Journal of the American Statistical Association 85, 813-823.

Anderson, T.W., 1951. Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions. Annals of Mathematical Statistics 22, 327-351.

Andrews, D.W.K., 1991. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. Econometrica 59, 817-858.

Andrews, D.W.K., Monahan, J.C., 1992. An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. Econometrica 60, 953-966.

Andrews, D.W.K., Ploberger, W., 1994. Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative. Econometrica 62, 1383-1414.

Becker, G.S., 1993. Human Capital, 3rd edition. University of Chicago Press, Chicago.

Bekker, P., Dobbelstein, P., Wansbeek, T., 1996. The APT Model as Reduced Rank Regression. Journal of Business and Economic Statistics 14, 199-202.

Blau, F.D., Kahn, L.M., 2000. Gender Differences in Pay. Journal of Economic Perspectives, Vol. 14 (4), 75-99.

Borjas, G.J., 2000. Labor Economics, 2nd edition. Irwin McGraw Hill.

Campbell, J.Y., 1987. Stock Returns and the Term Structure. Journal of Financial Economics 18, 373-399.

Chen, N., Roll, R., Ross, S.A., 1986. Economic forces and the stock market. Journal of Business 59, 383-403.

Cochrane, J.H., 2001. Asset Pricing. Princeton University Press, Princeton, NJ.

Costa, M., Gardini, A., Paruolo, P., 1997. A Reduced Rank Regression Approach to Tests of Asset Pricing. Oxford Bulletin of Economics and Statistics 59, 163-181.

den Haan, W., Levin, A., 1996. Inferences from parametric and non-parametric covariance matrix estimation procedures. Working paper, National Bureau of Economic Research.

den Haan, W., Levin, A., 1997. A Practitioner's Guide to Robust Covariance Matrix Estimation. Handbook of Statistics, Volume 15, G.S. Maddala and C.R. Rao (Eds.). Elsevier Science, North-Holland, Amsterdam.

Enders, W., 1995. Applied Econometric Time Series. John Wiley & Sons, Inc., New York.

Fama, E.F., French, K.R., 1992. The cross section of expected stock returns. The Journal of Finance 47, 427-465.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

Ferson, W.E., Foerster, S., 1994. Finite sample properties of the generalized method of moments in tests of conditional asset pricing models. Journal of Financial Economics 36, 29-55.

Gill, L., Lewbel, A., 1992. Testing the Rank and Definiteness of Estimated Matrices With Applications to Factor, State-Space and ARMA models. Journal of the American Statistical Association, Vol. 87, No. 419, Theory and Methods.

Gouriéroux, C., Monfort, A., 1989. A General Framework for Testing a Null Hypothesis in a "Mixed" Form. Econometric Theory 5, 63-82.

Gouriéroux, C., Monfort, A., Renault, E., 1993. Test sur le Noyau, l'Image et le Rang de la Matrice des Coefficients d'un Modle Linaire Multivari," Annales d'Economie et de Statistique, 32, 81-112.

Greene, W. H., 2000. Econometric Analysis, 4th edition. Prentice Hall, New Jersey.

Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14, 262-280.

Hausman, J.A., 1978. Specification Tests in Econometrics. Econometrica 46, 1251-1272.

Jaeger, D., Page, M., 1996. Degrees matter: New evidence on sheepskin effects in the returns to education. Review of Economics and Statistics 78, 733-740.

Johansen, S., 1988. Statistical Analysis of Cointegration Vectors. Journal of Economic Dynamics and Control 12, 231-254.

Johansen S., 1991. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. Econometrica 59, 1551-1580.

Kleinbergen, F., 1999. Reduced Rank Regression Using GMM. In: Generalized Method of Moments Estimation, L. Mátyás (Ed.). Cambridge University Press, Cambridge.

Kline, R.B., 1998. Structural Equation Modelling. Guilford Press, New York.

Newey, W., West, K., 1994. Automatic Lag Selection in Covariance Matrix Estimation. Review of Economic Studies 61, 631653.

Ruggles, S., Sobek M., et al., 1997. Integrated Public Use Microdata Series: Version 2.0. Historical Census Projects, University of Minnesota, Minneapolis.

Ruud, P.A., 2000. An Introduction to Classical Econometric Theory. Oxford University Press, Oxford.

Stock, J., Watson, M., 1988. Testing for Common Trends. Journal of the American Statistical Association 83, 1097-1107.

Schwert, G.W., 1983, Size and stock returns, and other empirical anomalies. Journal of Financial Economics 12, 3-12.

Szroeter, J., 1983. Generalized Wald Methods for Testing Nonlinear Implicit and Overidentifying Restrictions. Econometrica 51, 335-353.

Vahid, F., Engle, R.F., 1993. Common trends and Common Cycles. Journal of Applied Econometrics 8, 341-360.

TABLE 1: Rejection Rates, Szroeter Explicit Mixed-Form Test

Panel A: Rejection Under Reduced Rank Hypothesis

				errors $N(0, I)$		errors $N(0, \Upsilon)$	
g	q	K	n	OLS	SUR	OLS	SUR
2	1	3	150	0.08	0.09	0.01	0.07
			300	0.04	0.05	0.01	0.05
	1	5	150	0.11	0.12	0.01	0.07
			300	0.08	0.08	0.00	0.08
3	1	4	150	0.11	0.12	0.00	0.06
			300	0.08	0.08	0.00	0.06
	1	7	150	0.16	0.19	0.00	0.09
			300	0.09	0.10	0.00	0.08
3	2	4	150	0.06	0.07	0.01	0.04
			300	0.07	0.07	0.00	0.06
	2	7	150	0.10	0.12	0.01	0.07
			300	0.08	0.08	0.00	0.09

Panel B: Rejection Under Full Rank Hypothesis

				errors $N(0, I)$		errors $N(0, \Upsilon)$	
g	q	K	n	OLS	SUR	OLS	SUR
2	1	3	150	0.63	0.64	0.44	0.65
			300	0.94	0.95	0.86	0.96
	1	5	150	0.54	0.58	0.27	0.56
			300	0.92	0.92	0.70	0.93
3	1	4	150	0.82	0.84	0.50	0.82
			300	1.00	1.00	0.95	0.99
	1	7	150	0.72	0.76	0.26	0.78
			300	0.98	0.99	0.81	0.99
3	2	4	150	0.49	0.51	0.25	0.51
			300	0.91	0.92	0.71	0.92
	2	7	150	0.38	0.43	0.11	0.43
			300	0.82	0.83	0.45	0.84

TABLE 2: Rejection Rates, Szroeter Implicit Mixed-Form Test

Panel A: Rejection Under Reduced Rank Hypothesis

				errors $N(0, I)$		errors $N(0, \Upsilon)$	
g	q	K	n	OLS	SUR	OLS	SUR
2	1	3	150	0.04	0.05	0.00	0.04
			300	0.05	0.05	0.01	0.04
	1	5	150	0.04	0.05	0.00	0.04
			300	0.06	0.07	0.00	0.05
3	1	4	150	0.05	0.05	0.00	0.04
			300	0.06	0.07	0.00	0.04
	1	7	150	0.05	0.07	0.00	0.05
			300	0.07	0.08	0.00	0.04
3	2	4	150	0.02	0.02	0.00	0.02
			300	0.04	0.05	0.00	0.03
	2	7	150	0.02	0.02	0.00	0.02
			300	0.04	0.05	0.00	0.04

Panel B: Rejection Under Full Rank Hypothesis

				errors $N(0, I)$		errors $N(0, \Upsilon)$	
g	q	K	n	OLS	SUR	OLS	SUR
2	1	3	150	0.51	0.51	0.29	0.51
			300	0.92	0.92	0.81	0.94
	1	5	150	0.36	0.39	0.13	0.36
			300	0.82	0.82	0.57	0.87
3	1	4	150	0.70	0.72	0.32	0.68
			300	0.98	0.99	0.91	0.99
	1	7	150	0.52	0.58	0.12	0.58
			300	0.96	0.96	0.63	0.97
3	2	4	150	0.53	0.58	0.05	0.25
			300	0.96	0.96	0.63	0.97
	2	7	150	0.13	0.15	0.00	0.11
			300	0.60	0.63	0.15	0.62

TABLE 3: Comparison to Score Tests

Panel A: Rejection Under Reduced Rank Hypothesis

				Covariance Matrix Estimator				
g	q	K	test	White	NW	NW-P	A	AM
2	1	3	Szroeter, implicit	0.04	0.06	0.06	0.05	0.04
			Score	0.05	0.05	0.04	0.05	0.05
	1	5	Szroeter, implicit	0.06	0.10	0.10	0.07	0.07
			Score	0.06	0.02	0.02	0.05	0.05
3	1	4	Szroeter, implicit	0.05	0.17	0.17	0.12	0.05
			Score	0.05	0.01	0.00	0.07	0.08
	1	7	Szroeter, implicit	0.10	0.36	0.32	0.11	0.09
			Score	0.04	0.01	0.01	0.02	0.03
3	2	4	Szroeter, implicit	0.03	0.03	0.02	0.03	0.03
			Score	0.03	0.02	0.02	0.04	0.02
	2	7	Szroeter, implicit	0.04	0.07	0.07	0.06	0.04
			Score	0.04	0.02	0.00	0.04	0.03

Panel B: Rejection Under Full Rank Hypothesis

				Covariance Matrix Estimator				
g	q	K	test	White	NW	NW-P	A	AM
2	1	3	Szroeter, implicit	0.50	0.55	0.54	0.53	0.50
			Score	0.61	0.40	0.36	0.48	0.46
	1	5	Szroeter, implicit	0.42	0.48	0.47	0.41	0.40
			Score	0.47	0.19	0.13	0.30	0.30
3	1	4	Szroeter, implicit	0.69	0.82	0.82	0.75	0.73
			Score	0.75	0.22	0.15	0.44	0.54
	1	7	Szroeter, implicit	0.66	0.85	0.85	0.73	0.66
			Score	0.57	0.04	0.02	0.14	0.28
3	2	4	Szroeter, implicit	0.26	0.31	0.31	0.28	0.26
			Score	0.45	0.30	0.27	0.35	0.35
	2	7	Szroeter, implicit	0.16	0.27	0.28	0.19	0.16
			Score	0.31	0.08	0.05	0.15	0.18

TABLE 4: Tests of Other Models

Panel A: Regressor Values Differ Across Equations

	Covariance Matrix Estimator				
test	White	NW	NW-P	A	AM
	Reduced Rank				
Szroeter, explicit	0.05	0.06	0.05	0.05	0.05
Szroeter, implicit	0.07	0.08	0.08	0.07	0.06
Score	0.05	0.04	0.04	0.05	0.05
	Full Rank				
Szroeter, explicit	0.82	0.82	0.82	0.83	0.81
Szroeter, implicit	0.63	0.66	0.66	0.63	0.62
Score	0.60	0.44	0.44	0.58	0.60

Panel B: Co-integration

	Covariance Matrix Estimator				
test	White	NW	NW-P	A	AM
	Reduced Rank				
Szroeter, explicit	0.06	0.08	0.07	0.06	0.05
Szroeter, implicit	0.06	0.08	0.07	0.06	0.06
Score	0.05	0.05	0.05	0.05	0.05
	Full Rank				
Szroeter, explicit	0.59	0.62	0.62	0.61	0.58
Szroeter, implicit	0.59	0.62	0.62	0.61	0.58
Score	0.56	0.45	0.41	0.52	0.44

TABLE 5: Income, Gender and Education

Panel A: Personal Income (in \$) Distribution, U.S. Men and Women, Year = 1990

statistic	men	women
mean	24,568.79	11,472.01
std. dev.	27,818.91	13,321.25
1st quartile	7,743.50	2,248.75
median	18,000.00	7,257.00
3rd quartile	32,000.00	16,185.75
% sample ≤ 0	6.9%	15.3%
sample size	2105	2289

Panel B: Estimates of Mean Income, by Category

		education			
		Less Than High School	High School	Bachelor's	More Than Bachelor's
male	estimate	12,139.16	23,352.79	41,466.95	61,993.03
	s.e.	1,077.64	821.08	1,636.04	3328.23
female	estimate	6,292.02	11,341.61	18,001.59	32,471.75
	s.e.	535.07	369.76	846.35	2998.27

Panel C: Estimates and Tests of Reduced-Rank in Coefficient Matrix

The hypothesis is $H_0 : \beta_b = A\beta_a$, where β_a and β_b are the 4×1 vectors of mean income (by education level) for men and women, respectively, and A is a constant.

	Form in which Parameters are Estimated	
	Explicit Mixed Form	Implicit Mixed Form
estimate of A	0.489	0.478
s.e.	0.026	0.012
estimate of β_a	n.a.	(12329.90, 23423.00, 40760.02, 63099.14)
s.e.	n.a.	(776.39, 563.08, 1201.69, 2940.09)
test of H_0	0.179	1.806
p-value	0.98	0.63

TABLE 6: Asset Returns, Firm Size and Factor Pricing

Panel A: Summary Statistics, sample 1959:02-1999:12, annualized, percentages

	r_{SMALL}	r_{LARGE}	r_{VW}	r_{DEF}	r_{TERM}	g_{IP}	g_{CONS}	π_{UI}	g_{MON}
Mean	8.21	6.41	6.35	1.85	1.40	3.43	2.09	0.00	6.68
Std. Dev.	67.14	50.48	50.73	0.80	18.82	10.54	5.41	2.51	4.86
Skewness	-0.18	-0.40	-0.43	0.18	0.23	-0.10	-0.21	0.52	0.16
Kurtosis	7.32	5.22	5.31	2.26	7.06	9.03	4.49	7.61	5.15

Panel B: Unrestricted Estimates

The estimated model is: $y_{it} = \beta_i x_t^* + \varepsilon_{it}$, $i = 1, 2$, time period 1959:02-1999:12 where $y_{1t} = r_{SMALL}$, $y_{2t} = r_{LARGE}$, β_i is a (7×1) vector of coefficients, $x_t^* = ($ r_{VW}, r_{DEF}, r_{TERM}, g_{IP}, g_{CONS}, π_{UI}, $g_{MON})'$ and ε_{it} is the regression error.

	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{14}$	$\hat{\beta}_{15}$	$\hat{\beta}_{16}$	$\hat{\beta}_{17}$
est.	1.139	2.868	-0.238	-0.148	0.740	-0.627	-0.706
s.e.	(0.052)	(1.305)	(0.091)	(0.163)	(0.286)	(0.629)	(0.324)
	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$	$\hat{\beta}_{23}$	$\hat{\beta}_{24}$	$\hat{\beta}_{25}$	$\hat{\beta}_{26}$	$\hat{\beta}_{27}$
est.	0.994	0.027	0.015	0.008	-0.039	0.023	0.007
s.e.	(0.003)	(0.101)	(0.006)	(0.009)	(0.019)	(0.041)	(0.028)

Panel C: Tests of Reduced Rank

	Szroeter implicit mixed form	Szroeter explicit mixed form	score
stat	22.26	20.84	17.9
p-value	0.001	0.002	0.006

TABLE 6, Cont'd

Panel D: Tests for Residual Heteroscedasticity and Correlation

Residuals are calculated using OLS estimates, equation by equation ; Pearson = chi-square test for correlation; White test = F test with no cross terms; Q = Q statistic for testing 12 lags of autocorrelation; p-values in parentheses.

Residual Property		Size	Test	59-99
correlation	across equations		Pearson	-0.839 (0.000)
	across time	small	Q	78.047 (0.000)
		large	Q	42.333 (0.000)
heteroskedasticity		small	White	6.726 (0.000)
		large	White	4.482 (0.000)