

Generalized Reduced Rank Regression

Peter Reinhard Hansen¹

Working Paper No. 2002-02

January, 2002



Brown University

Department of Economics, Box B

Providence, RI 02912, USA

Phone: (401) 863 9864

Email: Peter_Hansen@brown.edu

¹I thank James D. Hamilton for many valuable comments and Jean-Marie Dufour, Jinyong Hahn, Paulo Paruolo, and Juan Toro for references. All errors remain my responsibility. Financial support from the Danish Research Agency and the Salomon Research Awards at Brown University is gratefully acknowledged.

Abstract

I introduce a technique to estimate parameters in regressions with reduced rank parameters in a general setting. The framework can handle a general class of parameter restrictions and allows for specifications with heteroskedastic and autocorrelated regression errors. Applications of this technique include: estimation of structural equations, estimation of reduced rank matrices in cross-section, panel, and time-series analysis, including estimation of cointegration relations in time series and panels.

JEL Classification: C3, C13, C31, C32, C33

Keywords: Estimation, Reduced Rank Regression, FIML, Panel-cointegration, Cointegration with Heteroskedasticity and Autocorrelation.

1 Introduction

In this paper I derive a new estimation technique called generalized reduced rank regression (GRRR) that can estimate the parameters in regression models in a general framework. The contribution of this paper is to link the GRRR framework to existing econometric problems and to provide a useful estimation technique to this class of regression problems. The novelty of the GRRR method is that it provides a unified framework for estimation of a general class of regression models by exploiting the structure of the estimation problem. The class includes models with: (i) reduced rank parameters, (ii) linear parameter restrictions, and (iii) a general covariance structure; and standard techniques such as OLS, GLS, RGLS, and reduced rank regression (RRR) are special (simplified) cases of the GRRR technique.

To illustrate the applicability of the GRRR technique, consider the regression problem $Y = X\pi' + Z\gamma' + \varepsilon$, where Y , X , Z , and ε have dimensions $T \times p$, $T \times p_1$, $T \times p_2$, and $T \times p$ respectively, where π and γ have dimensions $p \times p_1$ and $p \times p_2$ respectively, and where $\text{vec}(\varepsilon') \sim N_{Tp}(0, \Sigma)$. In special cases OLS or GLS may be applicable for the parameter estimation, whereas the GRRR method is useful in situations where the parameters may be subject to restrictions of the form: (i) $\pi = \alpha\beta'$ where α and β have dimensions $p \times r$ and $p_1 \times r$ respectively, (ii) $\text{vec}(\alpha, \gamma) = G\psi + g$ and $\text{vec}(\beta) = H\varphi + h$ for known matrices (vectors) H and G (h and g), and (iii) $\Sigma = \Sigma(\theta)$ where θ are the free parameters in the $Tp \times Tp$ covariance matrix, Σ .

Models with reduced rank parameters are common in econometrics. In some models the reduced rank parameters are a natural part of the model, in other models the reduced rank parameters are introduced in order to reduce the number of free parameters. Examples of the former include factor models, (see e.g., Bai and Serena, 1999), structural models, and cointegration models, (see e.g., Johansen, 1988) and recent examples of the latter include forecasting models (see e.g., Stock and Watson, 1999), instrumental variable estimation with many (potential weak) instruments, and GMM estimation (see

e.g., Hall and Inoue, 2001).

It is simple to impose a particular matrix to have reduced rank by expressing it as a product of two matrices, and the simplest estimation problem with reduced rank parameters is the RRR problem. RRR is closely related to the analysis of canonical covariates and canonical correlations of Hostelling (1935, 1936) and Anderson (1951, 1984). Discussion of these relations can be found in Anderson (1951), Izenman (1975), Tso (1981), Davies and Tso (1982), and Reinsel and Velu (1998). Reduced rank models with stationary variables have been analyzed by Velu, Reinsel, and Wichern (1986) and Velu and Reinsel (1987), Gudmundsson (1977) applied reduced rank regression to select linear combinations that best described the average variation in a vector of dependent variables. RRR is also used in the analysis of cointegrated variables. The concept of cointegration was introduced by Granger (1981), although the paper by Box and Tiao (1977) came close to introducing it.¹ Reduced rank regression was later applied to cointegration models by Johansen (1988) and Ahn and Reinsel (1990) in the vector autoregressive framework. In a Bayesian setting, reduced rank regression has been analyzed by Geweke (1996), and Costa, Gardini, and Paruolo (1997) applied reduced rank regression to test asset pricing models. The book by Reinsel and Velu (1998) provides an excellent exposition of RRR and its relation to several econometric estimation problems.

In a regression with unrestricted parameters, it is well known that minimizing the sum-of-squared residuals is equivalent to the method of maximum likelihood, if the errors in the regression equation are assumed to be iid Gaussian with an unknown covariance matrix. In this case, the parameter estimates are given from the least squares method. The estimation problem is only slightly more complicated when a reduced rank condition is imposed on a regression matrix. In this case the estimation problem

¹Box and Tiao applied canonical variates to decompose a process into linear combinations that are: white noise, stationary, and nearly non-stationary, and in relation to the nearly non-stationary variables, they argue against differencing the data prior to the analysis, due to the possible existence of “*stable contemporaneous linear relationships among the variables*” Box and Tiao (1977, p. 362).

can be solved as an eigenvalue problem, and this simplification is also applicable under a simple type of parameter restrictions. However, more complicated parameter restrictions often appear in econometrics and estimation currently requires the use of general estimation techniques that do not fully exploit the structure of the likelihood². The GRRR technique offers a better solution to this type of estimation problem.

Inference in a model with reduced rank parameters has (at least) three elements. One is to determine the rank(s), which affects the dimensions of several parameters, the second is the parameter estimation, and the third concerns the probabilistic properties of estimators. This paper contributes to the second problem – the estimation of parameters – whereas the two other aspects are not discussed in details. Determination of the rank(s) which defines the dimensions of several parameters as well as the probabilistic properties of the estimators, will depend on the data that are being analyzed. Fortunately, most of these results are readily available from the existing literature. For example, in a system with stationary variables, the (asymptotic) distribution of likelihood-based tests for the rank of a matrix of parameters are typically χ^2 , see Bartlett (1938) and Anderson (1951), whereas in systems with integrated variables the tests have a (squared) Dickey-Fuller type asymptotic distributions, see Johansen (1988).

The paper is organized as follows: Section 2 reviews and describes the reduced rank regression and estimation techniques for this problem. Section 3 presents the GRRR technique and Section 4 contains examples that show the applicability of the GRRR model. Section 5 contains some concluding remarks. Proofs are given in the Appendix. Bold font is used to denote matrices, $\text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_m)$ is used to denote the block-diagonal matrix that has the matrices $\mathbf{A}_1, \dots, \mathbf{A}_m$ along its diagonal and zeros in other entries, and $\mathbf{K}_{m,n}$ denotes the commutation matrix which is characterized by satisfying

²When general estimation techniques are applied to estimation of restricted reduced rank parameters, it is not unusual to experience slow convergence or even failure of the algorithm to converge to the global optimum.

the identity $\mathbf{K}_{m,n}\text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}')$ for any $m \times n$ matrix \mathbf{A} .

2 Reduced Rank Regression

Consider the linear regression equation:

$$Z_{0t} = \mathbf{\Pi}Z_{1t} + \varepsilon_t, \quad t = 1, \dots, T, \quad (1)$$

where Z_{0t} is a p -dimensional vector of dependent variables, Z_{1t} is a p_1 -dimensional vector of explanatory variables, $\{\varepsilon_t\}$ is a sequence of iid Gaussian variables, $\varepsilon_t \sim N(0, \mathbf{\Omega})$, and $\mathbf{\Pi}$ is an $p \times p_1$ matrix.

Reduced rank regression maximizes the Gaussian likelihood subject to the constraint $\text{rank}(\mathbf{\Pi}) = r$, where $0 \leq r \leq \min(p, p_1)$. The reduced rank of $\mathbf{\Pi}$ enables us to write $\mathbf{\Pi} = \mathbf{\alpha}\mathbf{\beta}'$, where $\mathbf{\alpha}$ is a $p \times r$ matrix and $\mathbf{\beta}$ is an $p_1 \times r$ matrix, both having full column rank. It is well known that the estimator for $\mathbf{\Pi}$ is given by the least squares estimator if $r = \min(p, p_1)$, whereas the estimator can be obtained by solving an eigenvalue problem whenever $r < \min(p, p_1)$. RRR can also be applied to estimate the parameters in the regression equation $Z_{0t} = \mathbf{\alpha}\mathbf{\beta}'Z_{1t} + \mathbf{\Psi}Z_{2t} + \varepsilon_t$, where we have p_2 additional regressors, Z_{2t} , as long as the $p \times p_2$ matrix $\mathbf{\Psi}$ is unrestricted. Johansen (1988) applied this method to estimate the parameters in the cointegrated vector autoregressive model. He showed that the estimation problem, for a given cointegration rank, simplifies to a reduced rank regression problem whereby he obtained the estimators. Johansen's technique is also applicable to estimation of parameters, which are subject to restrictions that take the form $\mathbf{\alpha} = \mathbf{G}\mathbf{\psi}$ and $\mathbf{\beta} = \mathbf{H}\mathbf{\varphi}$, for known matrices \mathbf{G} and \mathbf{H} , while problems of the form $\mathbf{\beta} = (\mathbf{H}_1\varphi_1, \dots, \mathbf{H}_r\varphi_r)$, for known matrices $\mathbf{H}_1, \dots, \mathbf{H}_r$, can be solved by a switching algorithm of Johansen and Juselius (1992), which reduces the estimation problem to a simple RRR problem in every iteration.³

³A similar decomposing of the estimation problem was made in Johansen and Swensen (1999) to test for rational expectation in the cointegrated VAR model.

Boswijk (1995) derived a more general estimation technique that solves estimation problems of the form $\text{vec}(\boldsymbol{\alpha}) = \mathbf{G}\psi$ and $\text{vec}(\boldsymbol{\beta}) = \mathbf{H}\varphi + h$, where $\text{vec}(\cdot)$ is the vectorization operator, \mathbf{G} and \mathbf{H} are known matrices, and h is a known vector.

When $\{\varepsilon_t\}$ is a sequence of iid Gaussian variables with mean zero and constant variance $\boldsymbol{\Omega}$ (a $p \times p$ matrix), the techniques yield the maximum likelihood estimators when Z_{1t} and Z_{2t} are measurable with respect to \mathcal{F}_{t-1} , where $\mathcal{F}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \dots)$.

The (GRRR) technique, derived below, can handle all of the estimation problems listed above, and the GRRR is applicable to a more general class of parameter restrictions, as well as models with a more complex structures of the covariance matrix, including heteroskedasticity and autocorrelation. In Section 4, I list a number of econometric models, where the GRRR technique is useful for estimation. The GRRR technique can be extended to non-linear restrictions by localized linear approximation, and thereby include the class of parameter restrictions considered by Elliott (1997, 2000), however, the non-linear aspect is not treated in this paper. The technique by Elliott (1997, 2000) uses minimum distance methods applied to the cointegrated regressions, see Engle and Granger (1987), whereas the technique in this paper is motivated by likelihood analysis.

3 Generalized Reduced Rank Regression

We define a generalized reduced rank regression as the regression

$$Z_{0t} = \boldsymbol{\alpha}\boldsymbol{\beta}'Z_{1t} + \boldsymbol{\Psi}Z_{2t} + \varepsilon_t, \quad t = 1, \dots, T, \quad (2)$$

where the parameters are subject to the restrictions

$$\begin{aligned} \text{vec}(\boldsymbol{\alpha}, \boldsymbol{\Psi}) &= \mathbf{G}\psi + g, \\ \text{vec}(\boldsymbol{\beta}) &= \mathbf{H}\varphi + h, \end{aligned}$$

for known matrices \mathbf{G} and \mathbf{H} and known vectors g and h . The errors are Gaussian distributed,

$$\varepsilon \equiv \text{vec}(\varepsilon_1, \dots, \varepsilon_T) \sim N_{Tp}(0, \Sigma).$$

The covariance matrix is either known or specified parametrically as $\Sigma = \Sigma(\theta)$, for some vector of parameters, θ . To simplify notation, we shall omit the argument “ θ ”, except where it is important to make the dependence of θ explicit. The three vectors of parameters, ψ , φ , and θ , are assumed to be variation free.⁴

It is the general class of restrictions and the general form of Σ , that makes the GRRR framework a powerful tool. Several estimation problems can be expressed in the form of a GRRR, and it is useful that one can specify Σ to accommodate heteroskedasticity and/or serial correlation, since an iid assumption is often implausible in economic applications.

If we set $\mathbf{G} = \mathbf{I}$, $\mathbf{H} = \mathbf{I}$, $g = 0$, $h = 0$, and $\Sigma = \mathbf{I}_T \otimes \Omega$, we see that the RRR is a special case of the GRRR.

A closer inspection of the estimation problem reveals a particular structure. When holding a subset of the parameters fixed, the partial estimation problem reduces to a restricted GLS problem, for which a closed-form solution is available. The GRRR estimation technique exploits this structure, by iterating on different subsets of the parameters. Since the likelihood function is increased in every iteration the value of the likelihood function is ensured to converge. Properties of estimation techniques that switch between subsets of the parameters has been analyzed by Oberhofer and Kmenta (1974).

Before we can present our main result, we need to introduce some additional notation. We define the matrices of observations, $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iT})'$, $i = 0, 1, 2$, and the

⁴This requires that the parameter space is a product space of the three parameter spaces of the three sets of parameters. So there is no functional dependence across parameters, which we need to take into account when estimating the parameters.

matrices:

$$\mathbf{Z}_{1\beta 2} = ((\mathbf{Z}_1\boldsymbol{\beta}, \mathbf{Z}_2) \otimes \mathbf{I}_p),$$

$$\mathbf{Z}_{1\alpha} = (\mathbf{Z}_1 \otimes \boldsymbol{\alpha})\mathbf{K}_{p_1, r},$$

where $\mathbf{K}_{p_1, r}$ is the commutation matrix and where \otimes denotes the Kronecker product.

Finally, define the matrices:

$$\mathbf{N}_{\beta, \Sigma} \equiv \mathbf{Z}'_{1\beta 2} \Sigma^{-1} \text{vec}(\mathbf{Z}'_0)',$$

$$\mathbf{M}_{\beta, \Sigma} \equiv \mathbf{Z}'_{1\beta 2} \Sigma^{-1} \mathbf{Z}_{1\beta 2},$$

$$\mathbf{N}_{\alpha, \Psi, \Sigma} \equiv \mathbf{Z}'_{1\alpha} \Sigma^{-1} \text{vec}(\mathbf{Z}'_0 - \Psi \mathbf{Z}'_2),$$

$$\mathbf{M}_{\alpha, \Sigma} \equiv \mathbf{Z}'_{1\alpha} \Sigma^{-1} \mathbf{Z}_{1\alpha}.$$

With these definitions, we can formulate the main result.

Theorem 1 *Let the parameter $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and Ψ be restricted by $\text{vec}(\boldsymbol{\alpha}, \Psi) = \mathbf{G}\boldsymbol{\psi} + g$ and $\text{vec}(\boldsymbol{\beta}) = \mathbf{H}\boldsymbol{\phi} + h$ and suppose that $\boldsymbol{\varepsilon} = \text{vec}(\varepsilon_1, \dots, \varepsilon_T) \sim N(0, \Sigma)$. Then the maximum likelihood estimators of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, Ψ , and Σ satisfy*

$$\text{vec}(\widehat{\boldsymbol{\alpha}}, \widehat{\Psi}) = \mathbf{G} \left[\mathbf{G}' \mathbf{M}_{\widehat{\boldsymbol{\beta}}, \widehat{\Sigma}} \mathbf{G} \right]^{-1} \mathbf{G}' \left(\mathbf{N}_{\widehat{\boldsymbol{\beta}}, \widehat{\Sigma}} - \mathbf{M}_{\widehat{\boldsymbol{\beta}}, \widehat{\Sigma}} g \right) + g, \quad (3)$$

$$\text{vec}(\widehat{\boldsymbol{\beta}}) = \mathbf{H} \left[\mathbf{H}' \mathbf{M}_{\widehat{\boldsymbol{\alpha}}, \widehat{\Sigma}} \mathbf{H} \right]^{-1} \mathbf{H}' \left(\mathbf{N}_{\widehat{\boldsymbol{\alpha}}, \widehat{\Psi}, \widehat{\Sigma}} - \mathbf{M}_{\widehat{\boldsymbol{\alpha}}, \widehat{\Sigma}} h \right) + h. \quad (4)$$

The maximum value of the likelihood function is given by

$$L_{\max}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\Psi}, \widehat{\theta}) = (2\pi)^{-\frac{Tp}{2}} |\widehat{\Sigma}|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} \widehat{\boldsymbol{\varepsilon}}' \widehat{\Sigma}^{-1} \widehat{\boldsymbol{\varepsilon}} \right),$$

where $\widehat{\boldsymbol{\varepsilon}} = \text{vec}(\widehat{\boldsymbol{\varepsilon}}_1, \dots, \widehat{\boldsymbol{\varepsilon}}_T)$, $\widehat{\boldsymbol{\varepsilon}}_t = Z_{0t} - \widehat{\boldsymbol{\alpha}} \widehat{\boldsymbol{\beta}}' Z_{1t} - \widehat{\Psi} Z_{2t}$, and where $\widehat{\Sigma} = \Sigma(\widehat{\theta})$.

In addition to the parameters α , β , and Ψ , we will have a likelihood equation for Σ ,

$$\hat{\theta} = \arg \max_{\theta} L(\hat{\alpha}, \hat{\beta}, \hat{\Psi}, \theta) \quad (5)$$

When (β, Σ) is fixed, the estimation problem for (α, Ψ) reduces to a restricted GLS problem and so does the estimation problem for β when (α, Ψ, Σ) . This is reflected in the structure of (3) and (4). The likelihood equation for Σ is given in its general form in (5). The exact form of (5) will depend on the specification chosen for Σ . Section 4 contains some special cases for which a closed-form expression of (5) can be obtained.

The definitions of $\mathbf{N}_{\beta, \Sigma}$, $\mathbf{M}_{\beta, \Sigma}$, $\mathbf{N}_{\alpha, \Psi, \Sigma}$, and $\mathbf{M}_{\alpha, \Sigma}$ involve the inverse of the $Tp \times Tp$ matrix Σ . If Tp is large it can be quite burdensome to work directly with the matrix Σ^{-1} . Fortunately, it is possible to simplify the expressions in some cases. If $\{\varepsilon_t\}$ is a sequence of independent variables the expressions can be simplified considerably. Lemma 2 covers the general case whereas the simplified expressions, achieved with additional assumptions, are given in Lemmas 3 and 4.

Lemma 2 *Let $\Sigma_{[t\tau]}^{-1}$ be the $p \times p$ sub-matrix of Σ^{-1} , such that the (i, j) th element of $\Sigma_{[t\tau]}^{-1}$ is the $(p(t-1) + i, p(\tau-1) + j)$ th element of Σ^{-1} .*

Then

$$\mathbf{N}_{\beta, \Sigma} = T^{-1} \sum_{t, \tau=1}^T \text{vec} \left(\Sigma_{[t\tau]}^{-1} Z_{0\tau} (Z'_{1t} \beta, Z'_{2t}) \right), \quad (6)$$

$$\mathbf{M}_{\beta, \Sigma} = T^{-1} \sum_{t, \tau=1}^T \left[\begin{pmatrix} \beta' Z_{1t} Z'_{1\tau} \beta & \beta' Z_{1t} Z'_{2\tau} \\ Z_{2t} Z'_{1\tau} \beta & Z_{2t} Z'_{2\tau} \end{pmatrix} \otimes \Sigma_{[t\tau]}^{-1} \right], \quad (7)$$

$$\mathbf{N}_{\alpha, \Psi, \Sigma} = T^{-1} \sum_{t, \tau=1}^T \text{vec} \left(Z_{1t} (Z_{0\tau} - \Psi Z_{2\tau})' \Sigma_{[t\tau]}^{-1} \alpha \right), \quad (8)$$

$$\mathbf{M}_{\alpha, \Sigma} = T^{-1} \sum_{t, \tau=1}^T \left[\alpha' \Sigma_{[t\tau]}^{-1} \alpha \otimes Z_{1t} Z'_{1\tau} \right]. \quad (9)$$

The advantage of Lemma 2 is that the equations do not involve matrices with a

dimension larger than $p \times p$. So with specification for Σ^{-1} rather than Σ , one can avoid the need to invert a $Tp \times Tp$ matrix.

In the following we consider situations where the structure of Σ leads to simplifications of these equations. In the absence of autocorrelation, Σ has a block diagonal structure, and we denote the T diagonal matrices of Σ by $\Omega(t)$, $t = 1, \dots, T$, and since $\Sigma_{[t\tau]}^{-1} = \Omega(t)^{-1}$ for $t = \tau$ and $\mathbf{0}$ otherwise, we obtain the simplifications of the following lemma.

Lemma 3 *Suppose that $\Sigma(\theta)$ is block diagonal, with T blocks of size $p \times p$ given by $\Omega(t)$, $t = 1, \dots, T$, then*

$$\begin{aligned} \mathbf{N}_{\beta, \Sigma} &= T^{-1} \sum_{t=1}^T \text{vec} \left(\Omega(t)^{-1} Z_{0t} (Z'_{1t} \beta, Z'_{2t}) \right), \\ \mathbf{M}_{\beta, \Sigma} &= T^{-1} \sum_{t=1}^T \left[\begin{pmatrix} \beta' Z_{1t} Z'_{1t} \beta & \beta' Z_{1t} Z'_{2t} \\ Z_{2t} Z'_{1t} \beta & Z_{2t} Z'_{2t} \end{pmatrix} \otimes \Omega(t)^{-1} \right], \\ \mathbf{N}_{\alpha, \Psi, \Sigma} &= T^{-1} \sum_{t=1}^T \text{vec} \left(Z_{1t} (Z_{0t} - \Psi Z_{2t})' \Omega(t)^{-1} \alpha \right), \\ \mathbf{M}_{\alpha, \Sigma} &= T^{-1} \sum_{t=1}^T [\alpha' \Omega(t)^{-1} \alpha \otimes Z_{1t} Z'_{1t}]. \end{aligned}$$

Under the assumptions of Lemma 3 the expression of the maximum value of the likelihood, see Theorem 1, simplifies to

$$L_{\max}(\hat{\alpha}, \hat{\beta}, \hat{\Psi}, \hat{\theta}) = (2\pi)^{-\frac{Tp}{2}} \left(\prod_{t=1}^T |\Omega(t)|^{-\frac{1}{2}} \right) \exp \left(-\frac{1}{2} \sum_{t=1}^T \hat{\varepsilon}'_t \Omega(t)^{-1} \hat{\varepsilon}_t \right).$$

Note that Lemma 3 allows for heteroskedasticity. If $\{\varepsilon_t\}$ is homoskedastic we obtain additional simplification. We define the moment matrices, $\mathbf{M}_{ij} \equiv T^{-1} \sum_{t=1}^T Z_{it} Z'_{jt}$, $i, j = 0, 1, 2$.

Lemma 4 *If $\text{var}(\varepsilon_t) = \Omega$, for all $t = 1, \dots, T$ and $\text{cov}(\varepsilon_t, \varepsilon_s) = \mathbf{0}$ for all $s \neq t$. We*

have the identities:

$$\begin{aligned}
\mathbf{N}_{\beta, \Sigma} &= \text{vec}(\mathbf{\Omega}^{-1}(\mathbf{M}_{01}\boldsymbol{\beta}, \mathbf{M}_{02})), \\
\mathbf{M}_{\beta, \Sigma} &= \left[\begin{array}{cc} \boldsymbol{\beta}'\mathbf{M}_{11}\boldsymbol{\beta} & \boldsymbol{\beta}'\mathbf{M}_{12} \\ \mathbf{M}_{21}\boldsymbol{\beta} & \mathbf{M}_{22} \end{array} \right] \otimes \mathbf{\Omega}^{-1}, \\
\mathbf{N}_{\alpha, \Psi, \Sigma} &= \text{vec}((\mathbf{M}_{10} - \mathbf{M}_{12}\boldsymbol{\Psi}')\mathbf{\Omega}^{-1}\boldsymbol{\alpha}), \\
\mathbf{M}_{\alpha, \Sigma} &= [\boldsymbol{\alpha}'\mathbf{\Omega}^{-1}\boldsymbol{\alpha} \otimes \mathbf{M}_{11}].
\end{aligned}$$

Under the iid assumptions the generic equation for the covariance parameters, (5), can be written as $\hat{\boldsymbol{\Omega}} = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'$, and the maximum value of the likelihood function simplifies to

$$L_{\max}^{-2/T}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Omega}}) = (2\pi)^{-p} |\hat{\boldsymbol{\Omega}}| \exp(p).$$

Some further simplification can be obtained under additional assumptions and in situations with less general restrictions, see Hansen (2000a, 2000b).

For a practical implementation of the GRRR technique one can first choose a set of initial parameter values, $(\boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)}, \boldsymbol{\Psi}^{(0)}, \theta^{(0)})$, and recursively update the parameter values using equations (3), (4), and (5) until the value of the likelihood function converges. Since the likelihood function is bounded by its maximum and the value of the likelihood function is increased in every iteration, the algorithm will eventually converge. However, it is not clear whether convergence to a local maximum or a saddle point can occur.⁵ In practice one can initialize the algorithm at different starting values and verify that the algorithm leads to the same value of the likelihood function.

⁵In simulation experiments where the GRRR algorithm was applied to the same data and initialized at different (randomly chosen) starting values, the algorithm always led to the same value of the likelihood. Matlab and Ox implementations of the GRRR technique is available from the author.

4 Applicability: Examples

Example 1 (FIML) Consider the structural equations expressed in the form

$$\mathbf{\Gamma}Y_t + \mathbf{\Phi}X_t = \eta_t, \quad t = 1, \dots, T,$$

where the matrices $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ have dimensions is a $m \times p$ and $m \times q$ respectively, where and η_t is iid Gaussian with mean zero and covariance matrix $\mathbf{\Lambda}$.

The reduced form of the model is given by

$$Y_t = \boldsymbol{\beta}'X_t + \varepsilon_t, \quad t = 1, \dots, T,$$

where $\boldsymbol{\beta}$ is a $q \times p$ matrix and ε_t is iid Gaussian with mean zero and covariance matrix $\mathbf{\Omega}$. The relation between the reduced form and the structural form are given by the well known identities, $\mathbf{\Gamma}\mathbf{\Omega}\mathbf{\Gamma}' = \mathbf{\Lambda}$ and $\mathbf{\Gamma}\boldsymbol{\beta}' + \mathbf{\Phi} = \mathbf{0}$.

We define

$$\boldsymbol{\alpha} = \mathbf{I}_p - \begin{pmatrix} \mathbf{\Gamma} \\ \mathbf{0}_{p-m \times p} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Psi} = \begin{pmatrix} -\mathbf{\Phi} \\ \mathbf{0}_{p-m \times q} \end{pmatrix},$$

where the zero matrices are added if the number of structural equations, m , is less than the number of endogenous variables, p . With these definitions it holds that

$$Y_t = \boldsymbol{\alpha}\boldsymbol{\beta}'X_t + \boldsymbol{\Psi}X_t + \varepsilon_t, \quad t = 1, \dots, T,$$

and since linear restrictions on $\text{vec}(\mathbf{\Gamma})$ and $\text{vec}(\mathbf{\Phi})$, can be formulated as linear restrictions on $\text{vec}(\boldsymbol{\alpha}, \boldsymbol{\Psi})$, the structural equations can be estimated using GRRR.

An advantage of the GRRR formulation, is that one can use restrictions on $\boldsymbol{\beta}$ for the identification of the structural parameters, as long as the restrictions do not violate the identity $\mathbf{\Gamma}\boldsymbol{\beta}' + \mathbf{\Phi} = \mathbf{0}$. On the other hand, the GRRR formulation does not allow

one to impose restrictions directly on the elements of $\mathbf{\Lambda} = \text{var}(\eta_t)$, but restrictions can be imposed on the reduced form covariance matrix, $\mathbf{\Omega}$.

The use of GRRR for estimation of structural models is similar to the estimation technique proposed by Hausman (1975). However, the GRRR allows for a more general specification of the covariance structure of ε_t (and hence η_t), such as heteroskedasticity and autocorrelation.

Example 2 (Moving average residuals) *The GRRR technique enables estimation of the cointegrated VARMA models. Since α , β , and Ψ are easily estimated for a fixed value of Σ , all we need to add is the likelihood equation for Σ . The general covariance matrix takes the form*

$$\Sigma = \begin{pmatrix} \mathbf{\Omega}_0 & \cdots & \mathbf{\Omega}'_q & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & & \ddots & \vdots \\ \mathbf{\Omega}_q & \cdots & \mathbf{\Omega}_0 & & & \mathbf{0} \\ & & & \ddots & & \\ \mathbf{0} & & & & \mathbf{\Omega}_0 & \cdots & \mathbf{\Omega}'_q \\ \vdots & \ddots & & & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & & \mathbf{\Omega}_q & \cdots & \mathbf{\Omega}_0 \end{pmatrix},$$

under the assumption that $\{\varepsilon_t\}$ is an MA(q) process, where $\mathbf{\Omega}_i = \text{cov}(\varepsilon_t, \varepsilon_{t-i})$, $i = 1, \dots, q$.

For fixed values of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\Psi}$, we calculate the residuals, $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_T$. These can be used to estimate Σ , by maximization of

$$\max_{\theta} L(\hat{\alpha}, \hat{\beta}, \hat{\Psi}, \theta) = (2\pi)^{-\frac{Tp}{2}} |\Sigma(\theta)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \hat{\varepsilon}' \Sigma(\theta)^{-1} \hat{\varepsilon}\right),$$

where θ contains the elements in $\mathbf{\Omega}_0, \dots, \mathbf{\Omega}_q$ (without duplication). There does not exist a closed form solution to this problem, but numeric methods are available, see Osborn

(1977) or Hamilton (1994). See Lütkepohl and Claessen (1997) for a different approach to estimation of cointegrated VARMA models.

Example 3 (ARCH-type heteroskedasticity) Let $\Omega(t)^{-1} = \mathbf{Q}'_t \mathbf{D} \mathbf{Q}_t$, where $\mathbf{Q}'_t = (\mathbf{I}_p, \mathbf{X}_t)$, \mathbf{X}_t is a sequence of $p \times q$ -dimensional exogenous variables, and \mathbf{D} is a $(p + q) \times (p + q)$ matrix of parameters. The parameter estimate of \mathbf{D} satisfies

$$\sum_{t=1}^T \mathbf{Q}_t (\mathbf{Q}'_t \widehat{\mathbf{D}} \mathbf{Q}_t)^{-1} \mathbf{Q}'_t = \sum_{t=1}^T \mathbf{Q}_t \widehat{\varepsilon}_t \widehat{\varepsilon}'_t \mathbf{Q}'_t. \quad (10)$$

This can be verified from the first order conditions. Let $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$, and $\widehat{\boldsymbol{\Psi}}$ be given, and define $(\widehat{\varepsilon}_1, \dots, \widehat{\varepsilon}_T)$ accordingly. The log-likelihood equation for \mathbf{D} is given by

$$l(\mathbf{D}) = \sum_{t=1}^T \log |\mathbf{Q}'_t \mathbf{D} \mathbf{Q}_t| - \text{tr} \left\{ \sum_{t=1}^T \widehat{\varepsilon}'_t \mathbf{Q}'_t \mathbf{D} \mathbf{Q}_t \widehat{\varepsilon}_t \right\},$$

and the differential is given by

$$\begin{aligned} l(\mathbf{D} + \mathbf{d}) &= \sum_{t=1}^T \log |\mathbf{Q}'_t (\mathbf{D} + \mathbf{d}) \mathbf{Q}_t| - \text{tr} \left\{ \sum_{t=1}^T \widehat{\varepsilon}'_t \mathbf{Q}'_t (\mathbf{D} + \mathbf{d}) \mathbf{Q}_t \widehat{\varepsilon}_t \right\} \\ &= l(\mathbf{D}) + \sum_{t=1}^T \text{tr} \left\{ (\mathbf{Q}'_t \mathbf{D} \mathbf{Q}_t)^{-1} \mathbf{Q}'_t \mathbf{d} \mathbf{Q}_t \right\} - \text{tr} \left\{ \sum_{t=1}^T \widehat{\varepsilon}'_t \mathbf{Q}'_t \mathbf{d} \mathbf{Q}_t \widehat{\varepsilon}_t \right\} + O(\|\mathbf{d}\|^2). \end{aligned}$$

So the first order condition is given by

$$\sum_{t=1}^T \text{tr} \left\{ [\mathbf{Q}_t (\mathbf{Q}'_t \mathbf{D} \mathbf{Q}_t)^{-1} \mathbf{Q}'_t - \mathbf{Q}_t \widehat{\varepsilon}_t \widehat{\varepsilon}'_t \mathbf{Q}'_t] \mathbf{d} \right\} = \mathbf{0}$$

for all $\mathbf{d} \in \mathbb{R}^{(p+q) \times (p+q)}$, which shows that (10) is the first order conditions for \mathbf{D} .

Example 4 (Structural change) Consider the cointegrated Gaussian VAR with a structural change in the cointegration relations and covariance matrix, see Hansen

(2000a). This can be expressed as

$$\Delta X_t = (\boldsymbol{\alpha}^o, \boldsymbol{\alpha}^o) \begin{pmatrix} \boldsymbol{\beta}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta}_2 \end{pmatrix}' \begin{pmatrix} X_{t-1} \mathbf{1}(t \leq T_1) \\ X_{t-1} \mathbf{1}(t > T_1) \end{pmatrix} + \boldsymbol{\Gamma}_1 \Delta X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T,$$

where $\mathbf{1}(\cdot)$ is the indicator function and where $\varepsilon_t \sim iid N(0, \boldsymbol{\Omega}_1)$ for $t \leq T_1$ and $\varepsilon_t \sim iid N(0, \boldsymbol{\Omega}_2)$ for $t > T_1$.

Set $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^o, \boldsymbol{\alpha}^o)$, $\boldsymbol{\beta} = \text{diag}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and $\boldsymbol{\Psi} = \boldsymbol{\Gamma}_1$. The regression problem can be written as a GRRR problem by the definitions $g = 0$, $h = 0$, and

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{pmatrix}, \quad \mathbf{G}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \mathbf{I}_{pr}, \quad \mathbf{G}_2 = \mathbf{I}_{p^2}, \quad \mathbf{H} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \otimes \mathbf{I}_{pr}$$

where p denotes the number of rows in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and r denotes the number of columns in $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$. The covariance matrix is given by $\boldsymbol{\Sigma} = \text{diag}(\mathbf{I}_{T_1} \otimes \boldsymbol{\Omega}_1, \mathbf{I}_{(T-T_1)} \otimes \boldsymbol{\Omega}_2)$ and the estimators are given by the sum of squares

$$\begin{aligned} \widehat{\boldsymbol{\Omega}}_1 &= T_1^{-1} \sum_{t=1}^{T_1} \hat{\varepsilon}_t \hat{\varepsilon}_t' \\ \widehat{\boldsymbol{\Omega}}_2 &= (T - T_1)^{-1} \sum_{t=T_1+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t', \end{aligned}$$

because the log-likelihood function splits into the sum

$$\begin{aligned} l(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) &\propto T_1 \log |\boldsymbol{\Omega}_1| + \text{tr} \left\{ \sum_{t=1}^{T_1} \hat{\varepsilon}_t \hat{\varepsilon}_t' \boldsymbol{\Omega}_1^{-1} \right\} \\ &\quad + (T - T_1) \log |\boldsymbol{\Omega}_2| + \text{tr} \left\{ \sum_{t=T_1+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t' \boldsymbol{\Omega}_2^{-1} \right\}, \end{aligned}$$

when $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\Psi}$ are taken as given.

Example 5 (Granger non-causality) Consider a cointegration model where X_{2t} does not Granger cause X_{1t} , (see Mosconi and Giannini (1992)). In the VAR(2) model, $\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \varepsilon_t$, for $t = 1, \dots, T$, the Granger non-causality is equivalent to the restrictions

$$\begin{aligned} \Pi &= \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \mathbf{0}_{p_2 \times p_1} & \Pi_{22} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \mathbf{0}_{p_2 \times r_1} & \alpha_{22} \end{pmatrix} \begin{pmatrix} \beta'_{11} & \beta'_{21} \\ \mathbf{0}_{r_2 \times p_1} & \beta'_{22} \end{pmatrix} \\ \Gamma_1 &= \begin{pmatrix} \Gamma_{1,11} & \Gamma_{1,12} \\ \mathbf{0}_{p_2 \times p_1} & \Gamma_{1,22} \end{pmatrix}, \end{aligned}$$

where p_1 is the dimension of X_{1t} , $p_2 = p - p_1$ is the dimension of X_{2t} , r_2 is the rank of Π_{22} , and $r_1 = r - r_2$. These restrictions are imposed using the matrices $\mathbf{G} = \text{diag}(\mathbf{G}_1, \mathbf{I}_{pr_2}, \mathbf{G}_3, \mathbf{I}_{p_2p})$ and $\mathbf{H} = \text{diag}(\mathbf{I}_{pr_1}, \mathbf{H}_2)$, where

$$\mathbf{G}_1 = \mathbf{I}_{r_1} \otimes \begin{pmatrix} \mathbf{I}_{p_1} \\ \mathbf{0}_{p_2} \end{pmatrix}, \quad \mathbf{G}_3 = \mathbf{I}_{p_1} \otimes \begin{pmatrix} \mathbf{I}_{p_1} \\ \mathbf{0}_{p_2} \end{pmatrix}, \quad \text{and } \mathbf{H}_2 = \mathbf{I}_{r_2} \otimes \begin{pmatrix} \mathbf{0}_{p_1} \\ \mathbf{I}_{p_2} \end{pmatrix}.$$

Example 6 (Panel cointegration) The panel cointegration model of Larsson, Lyhagen, and Lothgren (1998), Larsson and Lyhagen (1999), and Groen and Kleibergen (1999), take the form

$$\begin{pmatrix} \Delta X_{1t} \\ \vdots \\ \Delta X_{nt} \end{pmatrix} = \alpha \beta' \begin{pmatrix} X_{1,t-1} \\ \vdots \\ X_{n,t-1} \end{pmatrix} + \Gamma_1 \begin{pmatrix} \Delta X_{1,t-1} \\ \vdots \\ \Delta X_{n,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{pmatrix},$$

where α has one of three structures: $\alpha = \text{diag}(\alpha_1, \dots, \alpha_n)$, $\alpha = \mathbf{I}_n \otimes \alpha^o$, or α is unrestricted, where β has the structure: $\beta = \text{diag}(\beta_1, \dots, \beta_n)$, or $\beta = \mathbf{I}_n \otimes \beta^o$, where Γ_1 has the structure: $\Gamma_1 = \text{diag}(\Gamma_{1,1}, \dots, \Gamma_{1,n})$, $\Gamma_1 = \mathbf{I}_n \otimes \Gamma_1^o$, or Γ_1 is unrestricted, and where $\Sigma = \mathbf{I}_T \otimes \Omega$. The covariance matrix, Ω , can either be block diagonal $\Omega = \text{diag}(\Omega_1, \dots, \Omega_n)$ or be unrestricted.

Similar to the examples above, these (sets of) restrictions can be expressed in terms of \mathbf{G} and \mathbf{H} matrices.

Example 7 (Sector cointegration) Sector cointegration is similar to panel cointegration. The parameters have a block-diagonal structure except for one set of rows that corresponds to a common set of variables, X_{0t} . This can be expressed as

$$\begin{pmatrix} \Delta X_{0t} \\ \Delta X_{1t} \\ \vdots \\ \Delta X_{nt} \end{pmatrix} = \boldsymbol{\alpha}\boldsymbol{\beta}' \begin{pmatrix} X_{0,t-1} \\ X_{1,t-1} \\ \vdots \\ X_{n,t-1} \end{pmatrix} + \boldsymbol{\Gamma}_1 \begin{pmatrix} \Delta X_{0,t-1} \\ \Delta X_{1,t-1} \\ \vdots \\ \Delta X_{n,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{0t} \\ \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{pmatrix},$$

where $\boldsymbol{\alpha}$ may have the structure:

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_{01} & \cdots & \boldsymbol{\alpha}_{0n} \\ \boldsymbol{\alpha}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha}_n \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} (\boldsymbol{\alpha}_{01}, \dots, \boldsymbol{\alpha}_{0n}) \\ \mathbf{I}_n \otimes \boldsymbol{\alpha}^o \end{pmatrix},$$

or

$$\boldsymbol{\alpha} = \begin{pmatrix} (1, \dots, 1) \otimes \boldsymbol{\alpha}_0 \\ \mathbf{I}_n \otimes \boldsymbol{\alpha}^o \end{pmatrix},$$

and $\boldsymbol{\beta}$ may have the structure

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_{01} & \cdots & \boldsymbol{\beta}_{0n} \\ \boldsymbol{\beta}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\beta}_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} (\boldsymbol{\beta}_{01}, \dots, \boldsymbol{\beta}_{0n}) \\ \mathbf{I}_n \otimes \boldsymbol{\beta}^o \end{pmatrix},$$

or

$$\boldsymbol{\beta} = \begin{pmatrix} (1, \dots, 1) \otimes \boldsymbol{\beta}_0 \\ \mathbf{I}_n \otimes \boldsymbol{\beta}^o \end{pmatrix},$$

and a similar structure for $\boldsymbol{\Gamma}_1$.

5 Conclusion

We have derived an estimation technique that applies to parameter estimation of reduced rank parameters. The GRRR technique is applicable to a general class of estimation problems, which allows for a general specification of the covariance structure and where the parameter may be subject to restrictions. As illustrated by examples, the GRRR technique is applicable to several econometric models, including time-series, cross-section, and panel models. In the analysis of the cointegrated VAR the technique may prove particularly useful, as it allows for covariance specifications that include autocorrelation and heteroskedasticity, including multivariate ARCH-type specifications.

6 Acknowledgements

I thank James D. Hamilton for many valuable comments and Jean-Marie Dufour, Jinyong Hahn, Paulo Paruolo, and Juan Toro for references. All errors remain my responsibility. Financial support from the Danish Research Agency and the Salomon Research Awards at Brown University is gratefully acknowledged.

Appendix of Proofs

Proof of Theorem 1. We express (2) in the matrix notation

$$\mathbf{Z}_0 = \mathbf{Z}_1 \boldsymbol{\beta} \boldsymbol{\alpha}' + \mathbf{Z}_2 \boldsymbol{\Psi}' + \boldsymbol{\epsilon}, \tag{11}$$

where $\boldsymbol{\epsilon} = (\varepsilon_1, \dots, \varepsilon_T)'$, (note that $\text{vec}(\boldsymbol{\epsilon}) = \boldsymbol{\epsilon}$). The regression problem can be simplified to two partial regression problems, that take the restrictions into account individually.

Applying the vec operation to (11), see Magnus and Neudecker (1988), yields the equation

$$\begin{aligned} \text{vec}(\mathbf{Z}_0) &= (\mathbf{Z}'_1 \boldsymbol{\beta} \otimes \mathbf{I}_p) \text{vec}(\boldsymbol{\alpha}) + (\mathbf{Z}'_2 \otimes \mathbf{I}_p) \text{vec}(\boldsymbol{\Psi}) + \boldsymbol{\varepsilon} \\ &= [(\mathbf{Z}'_1 \boldsymbol{\beta}, \mathbf{Z}'_2) \otimes \mathbf{I}_p] \text{vec}(\boldsymbol{\alpha}, \boldsymbol{\Psi}) + \boldsymbol{\varepsilon} \\ &= \mathbf{Z}_{1\beta 2} (\mathbf{G}\boldsymbol{\psi} + g) + \boldsymbol{\varepsilon}, \end{aligned}$$

which may be rewritten as

$$\text{vec}(\mathbf{Z}_0) - \mathbf{Z}_{1\beta 2} g = \mathbf{Z}_{1\beta 2} \mathbf{G}\boldsymbol{\psi} + \boldsymbol{\varepsilon}.$$

For fixed values of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ this is a restricted GLS problem with the well know solution given by (3), where we have used the identities preceding Theorem 1.

Similarly, for fixed $\boldsymbol{\alpha}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Sigma}$, we have the equation

$$\begin{aligned} \text{vec}(\mathbf{Z}_0 - \boldsymbol{\Psi}\mathbf{Z}_2) &= \text{vec}(\boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Z}_1) + \boldsymbol{\varepsilon} \\ &= (\mathbf{Z}'_1 \otimes \boldsymbol{\alpha}) \text{vec}(\boldsymbol{\beta}') + \boldsymbol{\varepsilon} \\ &= (\mathbf{Z}'_1 \otimes \boldsymbol{\alpha}) \mathbf{K}_{p_1, r} \text{vec}(\boldsymbol{\beta}) + \boldsymbol{\varepsilon} \\ &= \mathbf{Z}_{1\alpha} \text{vec}(\boldsymbol{\beta}) + \boldsymbol{\varepsilon} \\ &= \mathbf{Z}_{1\alpha} (\mathbf{H}\boldsymbol{\phi} + h) + \boldsymbol{\varepsilon}, \end{aligned}$$

which we rewrite as

$$\text{vec}(\mathbf{Z}_0 - \boldsymbol{\Psi}\mathbf{Z}_2) - \mathbf{Z}_{1\alpha} h = \mathbf{Z}_{1\alpha} \mathbf{H}\boldsymbol{\phi} + \boldsymbol{\varepsilon}.$$

This is also a restricted GLS problem, with the solution given by (4). ■

Proof. of Lemmas 2, 3, and 4. The identity, (8), is proven by

$$\begin{aligned}
\mathbf{Z}'_{1\alpha}\Sigma^{-1}\text{vec}(\mathbf{Z}_0 - \Psi\mathbf{Z}_2) &= \mathbf{K}_{r,p_1} \sum_{t,\tau=1}^T (Z_{1t} \otimes \alpha') \Sigma_{[t\tau]}^{-1} \text{vec}(Z_{0\tau} - \Psi Z_{2\tau}) \\
&= \mathbf{K}_{r,p_1} \sum_{t,\tau=1}^T (Z_{1t} \otimes \alpha' \Sigma_{[t\tau]}^{-1}) \text{vec}(Z_{0\tau} - \Psi Z_{2\tau}) \\
&= \mathbf{K}_{r,p_1} \sum_{t,\tau=1}^T \text{vec}(\alpha' \Sigma_{[t\tau]}^{-1} (Z_{0\tau} - \Psi Z_{2\tau}) Z'_{1t}) \\
&= \sum_{t,\tau=1}^T \text{vec} \left(Z_{1t} (Z_{0\tau} - \Psi Z_{2\tau})' \Sigma_{[t\tau]}^{-1} \alpha \right),
\end{aligned}$$

and (9) is proven by

$$\begin{aligned}
\mathbf{Z}'_{1\alpha}\Sigma^{-1}\mathbf{Z}_{1\alpha} &= \mathbf{K}'_{p_1,r} (\mathbf{Z}_1 \otimes \alpha') \Sigma^{-1} (Z'_1 \otimes \alpha) \mathbf{K}_{p_1,r} \\
&= \mathbf{K}_{r,p_1} \sum_{t,\tau=1}^T (Z_{1t} \otimes \alpha') \Sigma_{[t\tau]}^{-1} (Z'_{1\tau} \otimes \alpha) \mathbf{K}_{p_1,r} \\
&= \mathbf{K}_{r,p_1} \sum_{t,\tau=1}^T (Z_{1t} \otimes \alpha' \Sigma_{[t\tau]}^{-1}) (Z'_{1\tau} \otimes \alpha) \mathbf{K}_{p_1,r} \\
&= \mathbf{K}_{r,p_1} \sum_{t,\tau=1}^T (Z_{1t} Z'_{1\tau} \otimes \alpha' \Sigma_{[t\tau]}^{-1} \alpha) \mathbf{K}_{p_1,r} \\
&= \sum_{t,\tau=1}^T (\alpha' \Sigma_{[t\tau]}^{-1} \alpha \otimes Z_{1t} Z'_{1\tau}).
\end{aligned}$$

The other identities, (6) and (7), are proven similarly. The eight identities in Lemmas 3 and 4 follow by setting $\Sigma_{[t\tau]}^{-1} = \mathbf{\Omega}(t)^{-1}$ or $\Sigma_{[t\tau]}^{-1} = \mathbf{\Omega}^{-1}$ for $t = \tau$ and zero otherwise. ■

References

- AHN, S. K., AND G. C. REINSEL (1990): “Estimation for Partially Non-Stationary 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.

- ANDERSON, T. W. (1984): *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley and Sons, 2nd edn.
- BAI, J., AND S. NG (1999): “Determining the Number of Factors in Approximate Factor Models,” *Boston College, Working Paper 440*.
- BARTLETT, M. S. (1938): “Further Aspects of the Theory of Multivariate Regression,” *Proceedings of the Cambridge Philosophical Society*, 34, 33–40.
- BOSWIJK, P. H. (1995): “Identifiability of Cointegrated Systems,” *Working paper, Tinbergen Institute*.
- BOX, G. E. P., AND G. C. TIAO (1977): “A Canonical Analysis of Multiple Time Series,” *Biometrika*, 64, 355–365.
- COSTA, M., A. GARDINI, AND P. PARUOLO (1997): “A Reduced Rank Regression Approach to Tests of Asset Pricing,” *Oxford Bulletin of economics and statistics*, 59, 163–181.
- DAVIES, P. T., AND M. K.-S. TSO (1982): “Procedures for Reduced-Rank Regression,” *Applied Statistics*, 31, 244–255.
- ELLIOTT, G. (1997): “Minimum Distance Methods and Cointegrated Models,” *UCSD manuscript*.
- (2000): “Estimating Restricted Cointegrating Vectors,” *Journal of Business and Economic Statistics*, 18, 91–99.
- ENGLE, R. F., AND C. W. J. GRANGER (1987): “Co-Integration and Error Correction: Representation, Estimation and Testing,” *Econometrica*, 55, 251–276.
- GEWEKE, J. (1996): “Bayesian Reduced Rank Regression in Econometrics,” *Journal of Econometrics*, 75, 121–146.
- GRANGER, C. W. J. (1981): “Some Properties of Time Series Data and their Use in Econometric Models Specification,” *Journal of Econometrics*, 16, 121–130.
- GROEN, J. J. J., AND F. R. KLEIBERGEN (1999): “Likelihood-Based Cointegration Analysis in Panels of Vector Error Correction Models,” *Tinbergen Institute Discussion Paper TI 99-055/4*.
- GUDMUNDSSON, G. (1977): “Multivariate Analysis of Economic Variables,” *Applied statistics*, 26, 48–59.
- HALL, A., AND A. INOUE (2001): “A Canonical Correlations Interpretation of GMM Estimation with Applications to Moment Selection,” *NCSU Mimeo*.
- HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton N.J.: Princeton University Press.
- HANSEN, P. R. (2000a): “Structural Changes in Cointegrated Processes,” Ph.D. thesis, University of California at San Diego.

- (2000b): “Structural Changes in the Cointegrated Vector Autoregressive Model,” *Brown University Economics Working Paper*.
- HAUSMAN, J. A. (1975): “An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models,” *Econometrica*, 43, 727–738.
- HOSTELLING, H. (1935): “The Most Predictable Criterion,” *Journal of education psychology*, 26, 139–142.
- (1936): “Relations Between Two Sets of Variables,” *Biometrika*, 28, 321–377.
- IZENMAN, A. J. (1975): “Reduced-Rank Regression for the Multivariate Linear Model,” *Journal of multivariate analysis*, 5, 248–264.
- JOHANSEN, S. (1988): “Statistical Analysis of Cointegration Vectors,” *Journal of Economic Dynamics and Control*, 12, 231–254.
- JOHANSEN, S., AND K. JUSELIOUS (1992): “Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK,” *Journal of Econometrics*, 53, 211–244.
- JOHANSEN, S., AND A. R. SWENSEN (1999): “Testing Exact Rational Expectations in Cointegrated Vector Autoregressive Models,” *Journal of Econometrics*, 93, 73–91.
- LARSSON, R., AND J. LYHAGEN (1999): “Likelihood-Based Inference in Multivariate Panel Cointegration Models,” *Working Paper Series in Economics and Finance, No. 331, Stockholm School of Economic*.
- LARSSON, R., J. LYHAGEN, AND M. LOTHGREN (1998): “Likelihood-Based Cointegration Test in Heterogeneous Panels,” *Working Paper Series in Economics and Finance No. 250, Stockholm School of Economics*.
- LÜTKEPOHL, H., AND H. CLAESSEN (1997): “Analysis of Cointegrated VARMA Processes,” *Journal of Econometrics*, 80, 223–239.
- MAGNUS, J. R., AND H. NEUDECKER (1988): *Matrix Differential Calculus with Applications in Statistics and Economics*. New York: Wiley.
- MOSCONI, R., AND C. GIANNINI (1992): “Non-Causality in Cointegrated Systems: Representation Estimation and Testing,” *Oxford Bulletin of Economics and Statistics*, 54, 399–417.
- OBERHOFER, W., AND J. KMENTA (1974): “A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models,” *Econometrica*, 42, 579–590.
- OSBORN, D. R. (1977): “Exact and Approximate Maximum Likelihood Estimators for Vector Moving Average Processes,” *Journal of Royal Statistical Society. Series B*, 39, 114–118.
- REINSEL, G. C., AND R. P. VELU (1998): *Multivariate Reduced-Rank Regression*. New York: Springer.

- STOCK, J., AND M. WATSON (1999): "Diffusion Indexes," *Working Paper*.
- TSO, M. K.-S. (1981): "Reduced-Rank Regression and Canonical Analysis," *Journal of Royal Statistical Society, series B*, 43, 183–189.
- VELU, R. P., AND G. C. REINSEL (1987): "Reduced Rank Regression with Autoregressive Errors," *Journal of Econometrics*, 35, 317–375.
- VELU, R. P., G. C. REINSEL, AND D. W. WICHERN (1986): "Reduced Rank Models for Multiple Time Series," *Biometrika*, 73, 105–118.