On the Estimation of Reduced Rank Regressions

Peter Reinhard Hansen[1]

Working Paper 2002-08

March, 2002

Brown University

Department of Economics

[1]Brown University, Department of Economics, Box B, Brown University, Providence, RI 02912, USA, Phone: (401) 863 9864, Email: Peter_Hansen@brown.edu

Abstract

It is well-know that estimation by reduced rank regression is given by the solution to a generalized eigenvalue problem. This paper presents a new proof to establish this result and provides additional insight into the structure of the estimation problem. The proof is a direct algebraic proof that some might find more intuitive than existing proofs.

JEL Classification: C3, C32

Keywords: Reduced Rank Regression; Least Squares Estimation

1 Introduction

Reduced rank regression (RRR) problems appear in several econometric models. Examples include the analysis of multivariate time-series, see Velu, Reinsel, and Wichern (1986) and Velu and Reinsel (1987) and the analysis of cointegrated variables in the vector autoregressive framework, see Johansen (1988, 1991, 1996). Reduced rank regression was introduced by Anderson (1951) and the book by Reinsel and Velu (1998) contains an excellent exposition of reduced rank regression and its relations to econometric models.

The objective in a RRR is to minimize the sum of squared residual subject to a reduced rank condition. Without the rank condition the estimation problem is a simple OLS problem. The properties RRR and OLS estimators have been analyzed and compared by Anderson (2002), in both a stationary and non-stationary setting. To show that the RRR estimators solve the minimization problem is not as simple as is the case for the OLS estimator. The estimation problem of a RRR can be simplified to the problem $\max_{x\in\mathbb{R}^{p\times r}} |x'Mx| / |x'Nx|$, where M and N are data-dependent matrices, and where $|A|$ denotes the determinant of a squared matrix A. The difficult step is to show that $\hat{x} = (\hat{v}_1, \ldots, \hat{v}_r)$ is the solution to this problem, where $\hat{v}_1, \ldots, \hat{v}_r$ are the the eigenvectors of $|\lambda N - M| = 0$ that corresponds to the r largest eigenvalues. This result can be obtained by a second order Taylor expansion, as in Johansen (1996); by reference to Poincare's theorem, see Magnus and Neudecker (1988); or by the algebraic proof presented in this paper. The new proof is based on a determinant representation that yields additional insight into estimation problems under additional restrictions.[1]

2 Reduced Rank Regression

A reduced rank regression takes the form

$$Z_{0t} = \alpha\beta' Z_{1,t} + \Psi Z_{2,t} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{1}$$

[1] Estimation of reduced rank parameters under additional parameter restrictions lead to more complicated estimation problems, which do not have closed-form solutions. Problems of this kind have been considered by Johansen and Juselius (1992), Boswijk (1995), and Hansen (2002) who proposed dexterous algorithms to solve the estimation problem.

where $Z_{0,t}$, $Z_{1,t}$, and $Z_{2,t}$ are vectors of dimension p, p_1, and p_2 respectively, and where α, β, and Φ are parameters of dimension $p \times r$, $p_1 \times r$, and $p \times p_2$ respectively. The error term, ε_t, is iid, with mean, $E(\varepsilon_t) = 0$, and variance $\mathrm{var}(\varepsilon_t) = \Omega$, and ε_t is independent of $(Z_{1t}, Z_{2,t}, Z_{1,t-1}, Z_{2,t-1}, \ldots)$. The RRR estimators of α, β, and Ψ are defined as the solution to $\min_{\alpha,\beta,\Psi} \left|\sum_{t=1}^{T} \varepsilon_t \varepsilon_t'\right|$, and the RRR estimator is the maximum likelihood estimator if ε_t is assumed to be normally distributed.

In matrix notation a RRR take the form $\mathbf{Z}_0 = \mathbf{Z}_1 \beta \alpha' + \mathbf{Z}_2 \Psi' + \boldsymbol{\varepsilon}$, where the tth row of $\mathbf{Z}_0$, $\mathbf{Z}_1$, $\mathbf{Z}_2$, and ε is given by $Z_{0,t}'$, $Z_{1,t}'$, $Z_{2,t}'$, and ε_t' respectively, $t = 1, \ldots, T$, so that $\mathrm{var}(\boldsymbol{\varepsilon}') = I_T \otimes \Omega$. We define the moment matrices $M_{ij} = \mathbf{Z}_i' \mathbf{Z}_j / T$, $i, j = 0, 1, 2$ and $S_{ij} = M_{ij} - M_{i2} M_{22}^{-1} M_{2i}$, $i, j = 0, 1$.

Theorem 1 (Reduced Rank Regression) *The parameter estimators of (1) are given by*

$$
\begin{aligned}
\hat{\beta} &= (\hat{v}_1, \ldots, \hat{v}_r)\phi, \\
\hat{\alpha}(\beta) &= S_{01}\hat{\beta}\left(\hat{\beta}' S_{11} \hat{\beta}\right)^{-1}, \\
\hat{\Psi} &= M_{02} M_{22}^{-1} - \hat{\alpha}\hat{\beta}' M_{12} M_{22}^{-1},
\end{aligned}
$$

where $(\hat{v}_1, \ldots, \hat{v}_r)$ are the eigenvectors corresponding to the r largest eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_r$ of $|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0$,[2] and where ϕ is an arbitrary $r \times r$ matrix with full rank.

Remark 1 *The parameters α and β are not identified. However, the $r \times r$ matrix, ϕ, can be used as a normalization device. E.g. if the normalization $\beta = (I_r, \beta_2')'$ is desired, one can choose ϕ to be the inverse of the matrix that consists of the first r rows of $(\hat{v}_1, \ldots, \hat{v}_r)$.*

Remark 2 *With a Gaussian likelihood, the MLE estimator for Ω is given by $\hat{\Omega} = S_{00} - S_{01}\hat{\beta}\left(\hat{\beta}' S_{11} \hat{\beta}\right)^{-1} \hat{\beta}' S_{10}$ and the maximum value of the likelihood is $L_{\max}^{-2/T}(\hat{\alpha}, \hat{\beta}, \hat{\Psi}, \hat{\Omega}) = (2\pi e)^p |S_{00}| \prod_{i=1}^{r}(1 - \hat{\lambda}_i)$.*

Remark 3 *Johansen (1988) applied RRR to the vector autoregressive model with cointegrated variables. The RRR structure appears from the error correction model, $\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t$, by setting $Z_{0t} = \Delta X_t$, $Z_{1t} = X_{t-1}$ and $Z_{2t} = (\Delta X_{t-1}', \ldots, \Delta X_{t-k+1}', D_t')'$.*

---

[2] *The eigenvectors satisfy $S_{10} S_{00}^{-1} S_{01} \hat{v}_i = \lambda_i S_{11} \hat{v}_i$, $\hat{v}_i' S_{11} \hat{v}_j = 1_{\{i=j\}}$. These are easily obtained using standard software such as Ox, Gauss, or Matlab, because $(\hat{v}_1, \ldots, \hat{v}_p) = S_{11}^{1/2}(x_1, \ldots, x_p)$, where $(x_1, \ldots, x_p)$ are the eigenvectors of the matrix $S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2}$. The eigenvalues of the two problems are the same.*

The following lemma is a central element of the proof of Theorem 1.

Lemma 2 *Let x be a $p \times r$ matrix, M and N be $p \times p$ symmetric matrices, where M is positive semi-definite and N is positive definite. Let $\lambda_1, \ldots, \lambda_p$ be the eigenvalues of $|\lambda N - M| = 0$, ordered in descending order, and let $v_1, \ldots, v_p$ be the corresponding eigenvectors.*

Then $\hat{x} = (v_1, \ldots, v_r)$ maximizes and $\tilde{x} = (v_{p-r+1}, \ldots, v_p)$ minimizes the function $f(x) = |x'Mx| \,/\, |x'Nx|$, and the maximum and minimum are given by $f(\hat{x}) = \prod_{i=1}^{r} \lambda_i$ and $f(\tilde{x}) = \prod_{i=p-r+1}^{p} \lambda_i$ respectively.

The proof of Johansen (1988) is based on a second order Taylor expansion of $\log f(x)$.[3] Below we shall present an algebraic proof, which applies a representation of determinants that involve products on non-squared matrices.

We introduce the following notation. Let $\mathbb{D}_p^r$ denote the set of all possible subsets of $J \subset \{1, \ldots, p\}$ that consist of $r \leq p$ distinct integers. For a given subset, $J \subset \mathbb{D}_p^r$, a $p \times r$ matrix, y, and a $p \times p$ matrix Λ, we define the $r \times r$ matrices $y_J = \{y_{ij}\}_{i \in J,\, j=1,\ldots,r}$ and $\Lambda_J = \{\Lambda_{ij}\}_{i,j \in J}$. We use $\text{diag}(a_1, \ldots, a_p)$ to denote the $p \times p$ diagonal matrix with diagonal elements: $a_1, \ldots, a_p$.

Example 1 *For $p = 3$, $r = 2$ we have $\mathbb{D}_3^2 = \{\{1,2\}, \{1,3\}, \{2,3\}\}$, and the subset $J = \{1, 2\}$ and the matrix*

$$y = \begin{pmatrix} y_{11} & y_{21} & y_{31} \\ y_{12} & y_{22} & y_{32} \end{pmatrix}', \quad \text{lead to } y_J = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix},$$

and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$ results $\Lambda_J = \text{diag}(\lambda_1, \lambda_2)$.

The following lemma provides a useful determinant representation.

Lemma 3 *Let $\Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$, and y a real $p \times r$ matrix, where $r \leq p$. Then with the definitions above, we have that*

$$|y'\Lambda y| = \sum_{J \in \mathbb{D}_p^r} |y_J' \Lambda_J y_J| = \sum_{J \in \mathbb{D}_p^r} |y_J' y_J| \, \Pi_{i \in J} \lambda_i = \sum_{J \in \mathbb{D}_p^r} |y_J|^2 \, \Pi_{i \in J} \lambda_i. \tag{2}$$

[3] The expression for the second order term is given in Johansen (1996), which corrects that in Johansen (1988).

Let $Q'\Lambda Q = S_{11}^{-1/2}S_{10}S_{00}^{-1}S_{01}S_{11}^{-1/2}$ be an orthogonal decomposition and define $y = QS_{11}^{1/2}\beta$. In the appendix we show that minimizing the determinant of the sum of squared residuals, is equivalent to maximizing $|y'\Lambda y|/|y'y|$, where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of $|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0$, and Lemma 3 shows that $|y'\Lambda y|/|y'y| = \sum_{J\in\mathbb{D}_p^r}|y_J|^2\,\Pi_{i\in J}\lambda_i$. So restrictions on β translate into restrictions on y through $y(\beta) = QS_{11}^{1/2}\beta$, which lead to restrictions on the possible convex combination, $\{|y_J|^2\}_{J\in\mathbb{D}_p^r}$, that one can take over the elements $\{\Pi_{i\in J}\lambda_i\}_{J\in\mathbb{D}_p^r}$. This observation may be useful for the estimation of reduced rank regressions that are subject to parameter restrictions on β, but we shall not attempt to address this issue in this paper.

3 Conclusion

This paper presented a new representation for determinants of products of non-squared matrices that led to a new algebraic proof of Theorem 1. The determinant representation provides additional insight into the estimation problem in the reduced rank regressions that are subject to parameter restrictions.

A Appendix

Proof of Lemma 3. The second and third equality follows trivially from $|AB| = |A|\,|B|$ for matrices of proper dimensions, and first holds trivially for $r = 1$ or $p = r$. So the proof is completed by induction as follows. Given that (i) (2) holds for $(p, r) = (\tilde{p} - 1, \tilde{r} - 1)$; and (ii) (2) holds for $(p, r) = (\tilde{p} - 1, \tilde{r})$, we show that (2) holds for $(p, r) = (\tilde{p}, \tilde{r})$. The following scheme

$p\backslash r$	1	2	3	4	$\cdots$
1	✓	–	–	–	
2	✓	✓	–	–	
3	✓	?	✓	–	
4	✓	?	?	✓	
$\vdots$	$\vdots$				$\ddots$

shows that this completes the proof.

Let $\tilde{\Lambda} \equiv \mathrm{diag}(\lambda_1, \ldots, \lambda_{p-1})$ and consider the case where $(y_{p1}, \ldots, y_{pr}) = (0, \ldots, 0)$. In this case we define $\tilde{y} \equiv \{y_{ij}\}_{i=1,\ldots,p-1}$, and by (ii) we have that

$$
\begin{aligned}
|y'\Lambda y| &= |\tilde{y}'\tilde{\Lambda}\tilde{y}| = \sum_{J\in\mathbb{D}^r_{p-1}} |y'_J y_J| \cdot \Pi_{i\in J}\lambda_i \\
&= \sum_{J\in\mathbb{D}^r_p, p\notin J} |y'_J y_J| \cdot \Pi_{i\in J}\lambda_i + \sum_{J\in\mathbb{D}^r_p, p\in J} |y'_J y_J| \cdot \Pi_{i\in J}\lambda_i.
\end{aligned}
$$

Since the last term is zero we have proven the lemma for the case where $(y_{p1},\ldots,y_{pr}) = 0$.

Assume now that $(y_{p1},\ldots,y_{pr}) \neq 0$. Choose a full rank $r \times r$-matrix Q, such that $(y_{p1},\ldots,y_{pr})Q = (0,\ldots,0,1)$ and define the $p-1 \times r-1$ matrix $\tilde{z}$ to be the first $r-1$ columns of $\tilde{y}Q$. Then it holds that

$$
|Q|^2\,|y'\Lambda y| = \left| Q'\tilde{y}'\tilde{\Lambda}\tilde{y}Q + \begin{pmatrix} 0_{r-1\times r-1} & 0 \\ 0 & \lambda_p \end{pmatrix} \right| = |Q'\tilde{y}'\tilde{\Lambda}\tilde{y}Q| + |\tilde{z}'\tilde{\Lambda}\tilde{z}|\lambda_p. \tag{3}
$$

By assumption (ii), the first term can be expressed as

$$
|Q'\tilde{y}'\tilde{\Lambda}\tilde{y}Q| = |Q|^2 \sum_{J\in\mathbb{D}^r_{p-1}} |\tilde{y}'_J\tilde{\Lambda}_J\tilde{y}_J| = |Q|^2 \sum_{J\in\mathbb{D}^r_p, p\notin J} |y'_J\Lambda_J y_J|. \tag{4}
$$

For $J \in \mathbb{D}^{r-1}_{p-1}$ we have that

$$
|\tilde{z}_J| = \left| \begin{pmatrix} \tilde{z}_J & 0 \\ 0 & 1 \end{pmatrix} \right| = |y_{\tilde{J}}Q|, \text{ and } \lambda_p|\tilde{\Lambda}_J| = |\Lambda_{\tilde{J}}|,
$$

where $\tilde{J} = \{J \cup \{p\}\} \in \mathbb{D}^r_p$, so the second term of (3) can be expressed as

$$
|\tilde{z}'\tilde{\Lambda}\tilde{z}|\lambda_p = \lambda_p = |Q|^2 \sum_{J\in\mathbb{D}^{r-1}_p, p\in J} |y'_J\Lambda_J y_J|, \tag{5}
$$

where we made use of assumption (i). Combining the identities (3–5), we have shown $|Q|^2|y'\Lambda y| = |Q|^2\sum_{J\in\mathbb{D}^r_p,p\notin J}|y'_J\Lambda_J y_J| + |Q|^2\sum_{J\in\mathbb{D}^r_p,p\in J}|y'_J\Lambda_J y_J| = |Q|^2\sum_{J\in\mathbb{D}^r_p}|y'_J\Lambda_J y_J|$, which completes the proof. ■

Lemma 4 *Let* $\Lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_p)$ *where* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ *and consider the function* $g(y) = |y'\Lambda y|\,/\,|y'y|$*. For* $\hat{y} = (I_r, 0_{r\times p-r})'$, $\tilde{y} = (0_{r\times p-r}, I_r)'$, *(the first* r *and last* r *unit vectors) it holds that* $\max_{y\in\mathbb{R}^{p\times r}} g(y) = g(\hat{y}) = \prod_{i=1}^r \lambda_i$ *and* $\min_{y\in\mathbb{R}^{p\times r}} g(y) = g(\tilde{y}) = \prod_{i=p-r+1}^p \lambda_i$.

Proof. By Lemma 3 we have that $g(y) = |y'\Lambda y| \,/\, |y'y| = \sum_{J\in\mathbb{D}_p^r} |y_J|^2 \,\Pi_{i\in J}\lambda_i / \sum_{J\in\mathbb{D}_p^r} |y_J|^2$, which is a convex combination over $\prod_{i\in J}\lambda_i$, $J \in \mathbb{D}_p^r$. Since the smallest and largest elements are $\prod_{i=1}^r \lambda_i$ and $\prod_{i=1}^r \lambda_i$, and these values can be obtained with $\tilde{y}$ and $\hat{y}$ the proof is complete. ■

Proof of Lemma 2. The matrix $(N^{-\frac{1}{2}}MN^{-\frac{1}{2}})$ is symmetric positive semi-definite, so we can diagonalize it as $N^{-\frac{1}{2}}MN^{-\frac{1}{2}} = Q'\Lambda Q$, where $Q'Q = I$, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$, and where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. By defining $y = QN^{\frac{1}{2}}x$, we have that $|x'Mx| \,/\, |x'Nx| = |y'\Lambda y| \,/\, |y'y|$. By Lemma 4, this is maximized (minimized) by $\hat{y} = (I_r, 0)'$ ($\tilde{y} = (0, I_r)'$), so $f(x)$ is maximized (minimized) by $\hat{x} = N^{-\frac{1}{2}}Q'\hat{y}$ ($\tilde{x} = N^{-\frac{1}{2}}Q'\tilde{y}$). ■

Proof of Theorem 1. The objective is to minimize $m_0(\alpha, \beta, \Psi)$, where

$$m_0(\alpha, \beta, \Psi) = \left| T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t' \right|, \qquad \hat{\varepsilon}_t = Z_{0t} - \alpha\beta' Z_{1t} - \Psi Z_{2t}.$$

It is simple to verify that $\arg\min_\Psi m(\alpha, \beta, \Psi) = \hat{\Psi}(\alpha, \beta) = M_{02}M_{22}^{-1} - \alpha\beta' M_{12}M_{22}^{-1}$, as it follows by the simple regression result. By defining the auxiliary variables, $R_{0t} = Z_{0t} - M_{02}M_{22}^{-1}Z_{2t}$ and $R_{1t} = Z_{1t} - M_{12}M_{22}^{-1}Z_{2t}$, the estimation problem is simplified to minimizing $m_1(\alpha, \beta) = |T^{-1}\sum_{t=1}^T \tilde{\varepsilon}_t\tilde{\varepsilon}_t'|$, where $\tilde{\varepsilon}_t = R_{0t} - \alpha\beta' R_{1t}$.

Similarly, we find that $\arg\min_\alpha m_1(\alpha, \beta) = \hat{\alpha}(\beta) = S_{01}\beta(\beta' S_{11}\beta)^{-1}$, and the simplified problem is now to minimize $m_2(\beta)$, where

$$\begin{aligned} m_2(\beta) &= T^{-1}\sum_{t=1}^{T}(R_{0t} - \hat{\alpha}(\beta)\beta' R_{1t})(R_{0t} - \hat{\alpha}(\beta)\beta' R_{1t})' \\ &= |S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}| = |S_{00}| \frac{|\beta'(S_{11} - S_{10}S_{00}^{-1}S_{01})\beta|}{|\beta' S_{11}\beta|}. \end{aligned}$$

Let $0 \leq \hat{\rho}_1 \leq \cdots \leq \hat{\rho}_p$ be the eigenvalues of $|\rho S_{11} - (S_{11} - S_{10}S_{00}^{-1}S_{01})| = 0$ and $\hat{v}_1, \ldots, \hat{v}_p$ the corresponding eigenvectors. Then, by Lemma 2, $\hat{\beta} = (v_1, \ldots, v_r)$ minimizes $m_2(\beta)$. The eigenvectors satisfy $[S_{11} - (S_{11} - S_{10}S_{00}^{-1}S_{01})]v_i = \rho_i v_i$, $i = 1, \ldots, p$. Since v_i is also an eigenvectors to $(S_{11} - S_{10}S_{00}^{-1}S_{01})$ with eigenvalue $\hat{\lambda}_i = 1 - \hat{\rho}_i$, it follows that the solution is given from the r largest eigenvalues of $|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0$. ■

Acknowledgements

Financial support from the Danish Research Agency and the Salomon Research Awards at Brown University is gratefully acknowledged.

References

Anderson, T. W. (1951): "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *Annals of Mathematical Statistics*, 22, 327–351.

——— (2002): "Reduced Rank Regression in Contegrated Models," *Journal of Econometrics*, 106, 203–216.

Boswijk, P. H. (1995): "Identifiability of Cointegrated Systems," *Working paper, Tinbergen Institute.*

Hansen, P. R. (2002): "Generalized Reduced Rank Regression," *Brown University, Economics Working Paper 2002-02.*

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

——— (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.

——— (1996): *Likelihood Based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press, Oxford, 2nd edn.

Johansen, S., and K. Juselius (1992): "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK," *Journal of Econometrics*, 53, 211–244.

Magnus, J. R., and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Economics.* Wiley, New York.

Reinsel, G. C., and R. P. Velu (1998): *Multivariate Reduced-Rank Regression.* Springer, New York.

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," *Biometrica*, 73, 105–118.