Johansen's Reduced Rank Estimator is GMM

Bruce E. Hansen*
University of Wisconsin†

January 2018

Abstract

The generalized method of moments (GMM) estimator of the reduced rank regression model is derived under the assumption of conditional homoskedasticity. We show that this GMM estimator is algebraically identical to the maximum likelihood estimator under normality developed by Johansen (1988). This includes the vector error correction model (VECM) of Engle and Granger. We also show that GMM tests for reduced rank (cointegration) are algebraically similar to the Gaussian likelihood ratio tests. This shows that normality is not necessary to motivate these estimators and tests.

*Research supported by the National Science Foundation.
†Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706.

1 Introduction

The vector error correction model (VECM) of Engle and Granger (1987) is one of the most widely used time-series models in empirical practice. The predominant estimation method for the VECM is the reduced rank regression method introduced by Johansen (1988, 1991, 1995). Johansen's estimation method is widely used because it is straightforward, a natural extension of the VAR model of Sims (1980), and is computationally tractable.

Johansen motivated his estimator as the maximum likelihood estimator (MLE) of the VECM under the assumption that the errors are i.i.d. normal. To this date there it is unclear if the estimator has a broader justification. In contrast, it is well known that least-squares estimation is both maximum likelihood under normality and method of moments under uncorrelatedness.

This paper provides the missing link. We show that Johansen's reduced rank estimator is algebraically identical to the generalized method of moments (GMM) estimator of the VECM, under the imposition of conditional homoskedasticity. This GMM estimator only uses uncorrelatedness and homoskedasticity. Thus Johansen's reduced rank estimator can be motivated under much broader conditions than normality.

The asymptotic efficiency of the estimator in the GMM class relies on the assumption of homoskedasticity (but not normality). When homoskedasticity fails the reduced rank estimator losses asymptotic efficiency, but retains its interpretation as a GMM estimator.

We also show that the GMM tests for reduced (cointegration) rank are nearly identical to Johansen's likelihood ratio tests. Thus the standard likelihood ratio tests for cointegration can be interpreted more broadly as GMM tests.

This paper does not introduce new estimation nor inference methods. It merely points out that the currently used methods have a broader interpretation than may have been understood. The results leave open the possibility that new GMM methods which do not impose homoskedasticity could be developed.

This paper is organized as follows. Section 2 introduces reduced rank regression models and Johansen's estimator. Section 3 presents the generalized method of moments and states the main theorems demonstrating equivalence of GMM and MLE. Section 4 presents the derivation of the GMM estimator. Section 5 contains two technical results relating generalized eigenvalue problems and the extrema of quadratic forms.

2 Reduced Rank Regression Models

The vector error correction model (VECM) for p variables of cointegrating rank r with k lags is

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + e_t \tag{1}$$

where D_t are the deterministic components. Observations are $t = 1, ..., T$. The matrices α and β are $p \times r$ with $r \leq p$. This is a famous workhorse model in applied time series, largely due to the seminal work of Engle and Granger (1987).

The primary estimation method for the VECM is known as reduced rank regression, and was developed Johansen (1988, 1991, 1995). Algebraically, the VECM (1) is a special case of the reduced rank regression model

$$Y_t = \alpha\beta' X_t + \Psi Z_t + e_t \tag{2}$$

where Y_t is $p \times 1$, X_t is $m \times 1$ and Z_t is $q \times 1$. The coefficient matrix α is $p \times r$ and β is $m \times r$ with $r \leq \min(m, p)$. Johansen derived the maximum likelihood estimator for model (2) under the assumption that e_t is i.i.d. $N(0, \Omega)$. This immediately applies to the VECM (1) and is the primary application of reduced rank regression in econometrics.

Reduced rank regression was first proposed by Anderson and Rubin (1949, 1950) and Anderson (1951). Indeed, these authors develop the maximum likelihood estimator for the model

$$Y_t = \Pi X_t + e_t \tag{3}$$

$$\Gamma'\Pi = 0 \tag{4}$$

where Γ is $p \times (p - r)$ and unknown. This is an alternative parameterization of (2) without the covariates Z_t. Anderson and Rubin (1949, 1950) considered the case $p - r = 1$ and primarily focus on estimation of the vector Γ. Anderson (1951) considered the case $p - r \geq 1$.

While the models (2) and (3)-(4) are equivalent and thus have the same MLE, the different parameterizations led the authors to different derivations. Anderson and Rubin derived the estimator of (3)-(4) by a tedious application of constrained optimization. (Specifically, the maximize the likelihood of model (3) imposing the constraint (4) using Lagrange multipler methods. The solution turns out to be tedious because (4) is a nonlinear function of the parameters Γ and Π.) The derivation is so cumbersome that it is excluded from nearly all statistics and econometrics textbooks, despite the fact that it is the source of the famous LIML estimator.

The elegant derivation used by Johansen (1988) is algebraically unrelated to that of Anderson-Rubin, and is based on applying a concentration argument to the product structure in (2). It is similar to the the derivation in Tso (1981) though the latter did not include the covariates Z_t. Johansen's derivation is algebraically straightforward and thus is widely taught to students.

Johansen's MLE for (2) is well known, but is stated here for completeness. Define the projection matrix $M_Z = I_T - Z(Z'Z)^{-1} Z'$, and the residual matrices $\widetilde{Y} = M_Z Y$ and $\widetilde{X} = M_Z X$. Consider the generalized eigenvalue problem

$$\left| \widetilde{X}'\widetilde{Y}\left(\widetilde{Y}'\widetilde{Y}\right)^{-1}\widetilde{Y}'\widetilde{X} - \widetilde{X}'\widetilde{X}\lambda \right| = 0. \tag{5}$$

Its solutions $1 > \widehat{\lambda}_1 > \cdots > \widehat{\lambda}_p > 0$ satisfy

$$\widetilde{X}'\widetilde{Y}\left(\widetilde{Y}'\widetilde{Y}\right)^{-1}\widetilde{Y}'\widetilde{X}\nu_i = \widetilde{X}'\widetilde{X}\widehat{\nu}_i\widehat{\lambda}_i.$$

$(\widehat{\lambda}_i, \widehat{\nu}_i)$ are known as the generalized eigenvalues and eigenvectors of $\widetilde{X}'\widetilde{Y}\left(\widetilde{Y}'\widetilde{Y}\right)^{-1}\widetilde{Y}'\widetilde{X}$ with respect to $\widetilde{X}'\widetilde{X}$. We impose the normalization $\widehat{\nu}_i'\widetilde{X}'\widetilde{X}\widehat{\nu}_i = 1$.

Given the normalization $\beta'\widetilde{X}'\widetilde{X}\beta = I_r$, Johansen's reduced rank estimator for β is

$$\widehat{\beta}_{\text{mle}} = \left[\widehat{\nu}_1, ..., \widehat{\nu}_r\right].$$

The MLE $\widehat{\alpha}_{\text{mle}}$ and $\widehat{\Psi}_{\text{mle}}$ are found by least-squares regression of Y_t on $\widehat{\beta}_{\text{mle}}'X_t$ and Z_t.

3 Generalized Method of Moments

Define $W_t = (X_t', Z_t')$. We derive the generalized method of moments (GMM) estimator of the reduced rank regression model (2) under the standard orthogonality restriction

$$\mathbb{E}\left(W_t e_t'\right) = 0$$

plus the homoskedasticity condition

$$\mathbb{E}\left(e_t e_t' \otimes W_t W_t'\right) = \Omega \otimes Q$$

where $\Omega = \mathbb{E}\left(e_t e_t'\right)$ and $Q = \mathbb{E}\left(W_t W_t'\right)$.

The efficient GMM criterion (see L. Hansen (1982)) takes the form

$$J_r(\alpha, \beta, \Psi) = T\overline{g}_r\left(\alpha, \beta, \Psi\right)' \widehat{V}^{-1}\overline{g}_r\left(\alpha, \beta, \Psi\right)$$

where

$$\overline{g}_r\left(\alpha, \beta, \Psi\right) = \frac{1}{T}\sum_{t=1}^{n}\left(\left(Y_t - \alpha\beta'X_t - \Psi Z_t\right) \otimes W_t\right) \tag{6}$$

$$\widehat{V} = \widehat{\Omega} \otimes \widehat{Q}$$

$$\widehat{\Omega} = \frac{1}{T}\sum_{t=1}^{n}\widehat{e}_t\widehat{e}_t' \tag{7}$$

$$\widehat{Q} = \frac{1}{T}\sum_{t=1}^{n}W_t W_t'$$

and $\widehat{e}_t$ are the least-squares residuals of the unconstrained model

$$\widehat{e}_t = Y_t - \widehat{\Pi}X_t - \widehat{\Psi}Z_t.$$

The GMM estimator are the parameters which jointly minimize the criterion $J_r(\alpha, \beta, \Psi)$ subject to a normalization for β. We use $\beta' \widetilde{X}' \widetilde{X} \beta = I_r$.

$$\left(\widehat{\alpha}_{\text{gmm}}, \widehat{\beta}_{\text{gmm}}, \widehat{\Psi}_{\text{gmm}}\right) = \underset{\beta' \widetilde{X}' \widetilde{X} \beta = I_r}{\text{argmin}} \; J_r(\alpha, \beta, \Psi).$$

The main contribution of the paper is the following surprising result.

Theorem 1. $\left(\widehat{\alpha}_{\text{gmm}}, \widehat{\beta}_{\text{gmm}}, \widehat{\Psi}_{\text{gmm}}\right) = \left(\widehat{\alpha}_{\text{mle}}, \widehat{\beta}_{\text{mle}}, \widehat{\Psi}_{\text{mle}}\right)$

Theorem 2. $J_r(\widehat{\alpha}_{\text{gmm}}, \widehat{\beta}_{\text{gmm}}, \widehat{\Psi}_{\text{gmm}}) = \text{tr}\left(\widehat{\Omega}^{-1}\left(\widetilde{Y}'\widetilde{Y}\right)\right) - Tp - T\sum_{i=1}^{r} \frac{\widehat{\lambda}_i}{1-\widehat{\lambda}_i}$ *where* $\widehat{\lambda}_i$ *are the eigenvalues from (5).*

Theorem 1 states that the GMM estimator is algebrically identical to the Gaussian maximum likelihood estimator.

This shows that Johansen's reduced rank regression estimator is not tied to the normality assumption. This is similar to the equivalence of least-squares as a method of moments estimator and the Gaussian MLE in the regression context.

The key is the use of the homoskedastic weight matrix. This shows that the Johansen reduced rank estimator is an efficient GMM estimator under conditional homoskedasticity. When homoskedasticity fails the Johansen reduced rank estimator continues to be a GMM estimator, but is no longer the efficient GMM estimator.

GMM hypothesis tests can be constructed by the difference in the GMM criteria. Consider tests for reduced rank, which in the context of VECM are tests for cointegration rank. Take the model

$$Y_t = \Pi X_t + \Psi Z_t + e_t$$

and consider hypotheses on reduced rank

$$\mathbb{H}_r : \text{rank}(\Pi) = r.$$

The GMM test statistic for $\mathbb{H}_r$ against $\mathbb{H}_{r+1}$ is

$$C_{r,r+1} = \min_{\beta' \widetilde{X}' \widetilde{X} \beta = I_r} J_r(\alpha, \beta, \Psi) - \min_{\beta' \widetilde{X}' \widetilde{X} \beta = I_{r+1}} J_{r+1}(\alpha, \beta, \Psi).$$

The GMM test statistic for $\mathbb{H}_r$ against $\mathbb{H}_p$ is

$$C_{r,p} = \min_{\beta' \widetilde{X}' \widetilde{X} \beta = I_r} J_r(\alpha, \beta, \Psi) - \min_{\beta' \widetilde{X}' \widetilde{X} \beta = I_p} J_p(\alpha, \beta, \Psi).$$

Theorem 3. *The GMM test statistics for reduced rank are*

$$C_{r,r+1} = T\left(\frac{\widehat{\lambda}_{r+1}}{1-\widehat{\lambda}_{r+1}}\right)$$
$$C_{r,p} = T\sum_{i=r+1}^{p}\frac{\widehat{\lambda}_i}{1-\widehat{\lambda}_i}$$

where $\widehat{\lambda}_i$ *are the eigenvalues from (5).*

Recall in contrast that the likelihood ratio test statistics derived by Johansen are

$$LR_{r,r+1} = -T\log\left(1-\widehat{\lambda}_{r+1}\right)$$
$$LR_{r,p} = -T\sum_{i=r+1}^{p}\log\left(1-\widehat{\lambda}_{r+1}\right).$$

The GMM test statistic $C_{r,r+1}$ and the LR statistic $LR_{r,r+1}$ yield equivalent tests since they are monotonic functions of one another. (If the bootstrap is used to assess significance, the two statistic will yield numerically identical p-values.) They are asymptotically identical under standard approximations, and in practice will be nearly identical since the eigenvalues $\widehat{\lambda}_i$ tend to be quite small in value so $-\log(1-\lambda) \approx \lambda/(1-\lambda) \approx \lambda$. For $p-(r+1) > 1$, the GMM test statistic $C_{r,p}$ and the LR statistic $LR_{r,p}$ are not equivalent tests (they cannot be written as monotonic functions of one another) but they also are asymptotically equivalent and will be nearly identical in practice.

4 Derivation of the GMM Estimator

It will be convenient to rewrite the criterion in standard matrix notation. Define the matrices Y, X, Z and W by stacking the observations. Model (2) is

$$Y = X\beta\alpha' + Z\Psi' + e.$$

The moment (6) is

$$\overline{g}_r(\alpha,\beta,\Psi) = \frac{1}{T}\operatorname{vec}\left(W'\left(Y - X\beta\alpha' - Z\Psi'\right)\right).$$

Using the relation

$$\operatorname{tr}(ABCD) = \operatorname{vec}\left(D'\right)'\left(C'\otimes A\right)\operatorname{vec}(A)$$

we obtain

$$\begin{aligned} J_r(\alpha,\beta,G) &= T\overline{g}_r(\alpha,\beta,\Psi)'\left(\widehat{\Omega}^{-1}\otimes\widehat{Q}^{-1}\right)\overline{g}_r(\alpha,\beta,\Psi) \\ &= \operatorname{vec}\left(W'\left(Y-X\beta\alpha'-Z\Psi'\right)\right)'\left(\widehat{\Omega}^{-1}\otimes\left(W'W\right)^{-1}\right)\operatorname{vec}\left(W'\left(Y-X\beta\alpha'-Z\Psi'\right)\right) \\ &= \operatorname{tr}\left(\widehat{\Omega}^{-1}\left(Y-X\beta\alpha'-Z\Psi'\right)'W\left(W'W\right)^{-1}W'\left(Y-X\beta\alpha'-Z\Psi'\right)\right). \end{aligned}$$

Following the concentration strategy used by Johansen, we fix β and concentrate out α and Ψ, obtaining a concentrated criterion which is a function of β only. The system is linear in the regressors $X\beta$ and Z. Given the homoskedastic weight matrix the GMM estimator of (α, Ψ) is multivariate least-squares. Using the partialling out (residual regression) approach, we can write the least-squares residual as the residual from the regression of $\widetilde{Y}$ on $\widetilde{X}\beta$, where $\widetilde{Y} = M_Z Y$ and $\widetilde{X} = M_Z X$ are the residuals from regressions on Z. That is, the least-squares residual is

$$\begin{aligned}\widehat{e}(\beta) &= \widetilde{Y} - \widetilde{X}\beta\left(\beta'\widetilde{X}'\widetilde{X}\beta\right)^{-1}\beta'\widetilde{X}'\widetilde{Y} \\ &= \widetilde{Y} - \widetilde{X}\beta\beta'\widetilde{X}'\widetilde{Y}\end{aligned}$$

where the second equality uses the normalization $\beta'\widetilde{X}'\widetilde{X}\beta = I_r$. Since the space spanned by $W = (X, Z)$ equals that spanned by $(\widetilde{X}, Z)$ we can write

$$W\left(W'W\right)^{-1}W' = Z\left(Z'Z\right)^{-1}Z' + \widetilde{X}\left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'.$$

Since $Z'\widehat{e}(\beta) = 0$ we find

$$\begin{aligned}W\left(W'W\right)^{-1}W'\widehat{e}(\beta) &= \widetilde{X}\left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'\widehat{e}(\beta) \\ &= \widetilde{X}\left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'\widetilde{Y} - \widetilde{X}\beta\beta'\widetilde{X}'\widetilde{Y}\end{aligned}$$

and

$$\begin{aligned}\widehat{e}(\beta)'W\left(W'W\right)^{-1}W'\widehat{e}(\beta) &= \widetilde{Y}'\widetilde{X}\left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'\widetilde{Y} - \widetilde{Y}'\widetilde{X}\beta\beta'\widetilde{X}'\widetilde{Y} \\ &= \widetilde{Y}'\widetilde{Y} - \widetilde{Y}'M_{\widetilde{X}}\widetilde{Y} - \widetilde{Y}'\widetilde{X}\beta\beta'\widetilde{X}'\widetilde{Y}\end{aligned}$$

where

$$M_{\widetilde{X}} = M_{\widetilde{X}} = I - \widetilde{X}\left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'.$$

Using the partialling out (residual regression) approach we can write the variance estimator (7) as

$$\widehat{\Omega} = \frac{1}{T}Y'\left(I - W\left(W'W\right)^{-1}W'\right)Y = \frac{1}{T}\widetilde{Y}'M_{\widetilde{X}}\widetilde{Y}.$$

Thus the concentrated GMM criterion is

$$\begin{aligned}J_r^*(\beta) &= \operatorname{tr}\left(\widehat{\Omega}^{-1}\widehat{e}(\beta)'W\left(W'W\right)^{-1}W'\widehat{e}(\beta)\right) \\ &= \operatorname{tr}\left(\widehat{\Omega}^{-1}\left(\widetilde{Y}'\widetilde{Y}\right)\right) - \operatorname{tr}\left(\widehat{\Omega}^{-1}\left(\widetilde{Y}'M_{\widetilde{X}}\widetilde{Y}\right)\right) - \operatorname{tr}\left(\widehat{\Omega}^{-1}\left(\widetilde{Y}'\widetilde{X}\beta\beta'\widetilde{X}'\widetilde{Y}\right)\right) \\ &= \operatorname{tr}\left(\widehat{\Omega}^{-1}\left(\widetilde{Y}'\widetilde{Y}\right)\right) - Tp - T\operatorname{tr}\left(\beta'\widetilde{X}'\widetilde{Y}\left(\widetilde{Y}'M_{\widetilde{X}}\widetilde{Y}\right)^{-1}\widetilde{Y}'\widetilde{X}\beta\right). \qquad (8)\end{aligned}$$

The GMM estimator minimizes $J_r^*(\beta)$, or equivalently maximizes the third term in (8). This is a generalized eigenvalue problem. Lemma 2 (in the next section) shows that the solution is $\widehat{\beta}_{\text{gmm}} = [\widetilde{\nu}_1, ..., \widetilde{\nu}_r]$ as claimed.

Since the estimates $\widehat{\alpha}_{\text{gmm}}$, $\widehat{\Psi}_{\text{gmm}}$ are found by regression given $\widehat{\beta}_{\text{gmm}}$, and this is equivalent with MLE, we also conclude that $\widehat{\alpha}_{\text{gmm}} = \widehat{\alpha}_{\text{mle}}$ and $\widehat{\Psi}_{\text{gmm}} = \widehat{\Psi}_{\text{mle}}$. This completes the proof of Theorem 1.

To establish Theorem 2, Lemma 2 also shows that the minimum of the criterion is

$$\begin{aligned}
J_r(\widehat{\alpha}_{\text{gmm}}, \widehat{\beta}_{\text{gmm}}, \widehat{\Psi}_{\text{gmm}}) &= \min_{\beta'\widetilde{X}'\widetilde{X}\beta = I_r} J_r(\alpha, \beta, G) \\
&= \min_{\beta'\widetilde{X}'\widetilde{X}\beta = I_r} J_r^*(\beta) \\
&= \operatorname{tr}\left(\widehat{\Omega}^{-1}\left(\widetilde{Y}'\widetilde{Y}\right)\right) - Tp - T \max_{\beta'\widetilde{X}'\widetilde{X}\beta = I_r} \operatorname{tr}\left(\beta'\widetilde{X}'\widetilde{Y}\left(\widetilde{Y}'M_{\widetilde{X}}\widetilde{Y}\right)^{-1}\widetilde{Y}'\widetilde{X}\beta\right) \\
&= \operatorname{tr}\left(\widehat{\Omega}^{-1}\left(\widetilde{Y}'\widetilde{Y}\right)\right) - Tp - T\sum_{i=1}^{r} \frac{\widehat{\lambda}_i}{1-\widehat{\lambda}_i}.
\end{aligned}$$

This establishes Theorem 2.

5 Extrema of Quadratic Forms

To establish Theorems 1 and 2 we need a simple extrema property. We first state a simple property which relates the maximization of quadratic forms to generalized eigenvalues and eigenvectors. It is a slight extension of Theorem 11.13 of Magnus and Neudecker (1988).

Lemma 1. *Suppose A and C are $p \times p$ real symmetric matrices with $C > 0$. Let $\lambda_1 > \cdots > \lambda_p > 0$ be the generalized eigenvalues of A with respect to C and $\nu_1, ..., v_p$ be the associated eigenvectors. Then*

$$\max_{\beta' C\beta = I_r} \operatorname{tr}\left(\beta' A \beta\right) = \sum_{i=1}^{r} \lambda_i$$

and

$$\underset{\beta' C\beta = I_r}{\operatorname{argmax}} \operatorname{tr}\left(\beta' A \beta\right) = [\nu_1, ..., \nu_r].$$

Proof. Define $\gamma = C^{1/2\prime}\beta$ and $\overline{A} = C^{-1/2}AC^{-1/2\prime}$. The eigenvalues of $\overline{A}$ are equal to the generalized eigenvalues λ_i of A with respect to B. The associated eigenvectors of $\overline{A}$ are $C^{1/2\prime}\nu_i$. Thus by Theorem 11.13 of Magnus and Neudecker (1988)

$$\max_{\beta' C\beta = I_r} \operatorname{tr}\left(\beta' A \beta\right) = \max_{\gamma'\gamma = I_r} \operatorname{tr}\left(\gamma'\overline{A}\gamma\right) = \sum_{i=1}^{r} \lambda_i$$

and

$$\underset{\beta' C\beta = I_r}{\operatorname{argmax}} \operatorname{tr}\left(\beta' A\beta\right) = C^{-1/2\prime} \underset{\gamma'\gamma = I_r}{\operatorname{argmax}} \operatorname{tr}\left(\gamma'\overline{A}\gamma\right)$$
$$= C^{-1/2\prime} C^{1/2\prime} [\nu_1, ..., \nu_r]$$
$$= [\nu_1, ..., \nu_r]$$

as claimed.

Lemma 2. *Let $M_X = I - X\left(X'X\right)^{-1} X'$. If $X'X > 0$ and $Y'M_XY > 0$ then*

$$\max_{\beta' X' X\beta = I_r} \operatorname{tr}\left(\beta' X'Y(Y'M_XY)^{-1}Y'X\beta\right) = \sum_{i=1}^{r} \frac{\lambda_i}{1-\lambda_i}$$

and

$$\underset{\beta' X' X\beta = I_r}{\operatorname{argmax}} \operatorname{tr}\left(\beta' X'Y(Y'M_XY)^{-1}Y'X\beta\right) = [\nu_1, ..., \nu_r]$$

where $1 > \lambda_1 > \cdots > \lambda_p > 0$ are the generalized eigenvalues of $X'Y(Y'Y)^{-1}Y'X$ with respect to $X'X$, and $\nu_1, ..., \upsilon_p$ are the associated eigenvectors.

Proof. By Lemma 1,

$$\max_{\beta' X' X\beta = I_r} \operatorname{tr}\left(\beta' X'Y(Y'M_XY)^{-1}Y'X\beta\right) = \sum_{i=1}^{r} \widetilde{\lambda}_i$$

and

$$\underset{\beta' X' X\beta = I_r}{\operatorname{argmax}} \operatorname{tr}\left(\beta' X'Y(Y'M_XY)^{-1}Y'X\beta\right) = [\widetilde{\nu}_1, ..., \widetilde{\nu}_r]$$

where $\widetilde{\lambda}_1 > \cdots > \widetilde{\lambda}_p > 0$ are the generalized eigenvalues of $X'Y(Y'M_XY)^{-1}Y'X$ with respect to $X'X$, and $\widetilde{\nu}_1, ..., \widetilde{\nu}_p$ are the associated eigenvectors. The proof is established by showing that $\widetilde{\lambda}_i = \lambda_i/(1-\lambda_i)$ and $\widetilde{\nu}_i = \nu_i$.

Let $(\widetilde{\nu}, \widetilde{\lambda})$ be a generalized eigenvector/eigenvalue pair of $X'Y(Y'M_XY)^{-1}Y'X$ with respect to $X'X$. The pair satisfies

$$X'Y\left(Y'M_XY\right)^{-1} Y'X\widetilde{\nu} = X'X\widetilde{\nu}\widetilde{\lambda}. \tag{9}$$

By the Woodbury matrix identity

$$\left(Y'M_XY\right)^{-1} = \left(Y'Y - Y'X\left(X'X\right)^{-1} X'Y\right)^{-1}$$
$$= \left(Y'Y\right)^{-1} + \left(Y'Y\right)^{-1} Y'X\left(X'X - X'Y\left(Y'Y\right)^{-1} Y'X\right)^{-1} X'Y\left(Y'Y\right)^{-1}$$
$$= \left(Y'Y\right)^{-1} + \left(Y'Y\right)^{-1} Y'X\left(X'M_YX\right)^{-1} X'Y\left(Y'Y\right)^{-1}$$

where $M_Y = I - Y\left(Y'Y\right)^{-1}Y'$. Thus

$$\begin{aligned}X'Y\left(Y'M_XY\right)^{-1}Y'X &= X'Y\left(Y'Y\right)^{-1}Y'X + X'Y\left(Y'Y\right)^{-1}Y'X\left(X'M_YX\right)^{-1}X'Y\left(Y'Y\right)^{-1}Y'X \\ &= X'P_YX + X'P_YX\left(X'M_YX\right)^{-1}X'P_YX \\ &= X'X\left(X'M_YX\right)^{-1}X'P_YX\end{aligned}$$

where $P_Y = Y\left(Y'Y\right)^{-1}Y'$ and the final equality uses $X'P_YX = X'X - X'M_YX$. Substituted into (9) we obtain

$$X'X\left(X'M_YX\right)^{-1}X'P_YX\widetilde{\nu} = X'X\widetilde{\nu}\widetilde{\lambda}.$$

Multiplying both sides by $\left(X'M_YX\right)\left(X'X\right)^{-1}$ this implies

$$\begin{aligned}X'P_YX\widetilde{\nu} &= X'M_YX\widetilde{\nu}\widetilde{\lambda} \\ &= X'X\widetilde{\nu}\widetilde{\lambda} - X'P_YX\widetilde{\nu}\widetilde{\lambda}.\end{aligned}$$

Collecting terms

$$X'P_YX\widetilde{\nu}(1+\widetilde{\lambda}) = X'X\widetilde{\nu}\widetilde{\lambda}$$

which implies

$$X'P_YX\widetilde{\nu} = X'X\widetilde{\nu}\frac{\widetilde{\lambda}}{(1+\widetilde{\lambda})}.$$

This is an eigenvalue equation. It shows that $\widetilde{\lambda}/(1+\widetilde{\lambda}) = \lambda$ is a generalized eigenvalue and $\widetilde{\nu}$ the associated eigenvector of $X'P_YX$ respect to $X'X$. Solving, $\widetilde{\lambda} = \lambda/(1-\lambda)$. This means that the generalized eigenvalues of $X'Y(Y'M_XY)^{-1}Y'X$ with respect to $X'X$ are $\lambda_i/(1-\lambda_i)$ and ν_i. Since $\lambda/(1-\lambda)$ is monotonically increasing on $[0,1)$ and $\lambda_i < 1$, it follows that the ordering of λ_i and $\widetilde{\lambda}_i$ are identical. Thus $\widetilde{\lambda}_i = \lambda_i/(1-\lambda_i)$ as claimed.

References

[1] Anderson, T. W. (1951): "Estimating linear restrictions on regression coefficeints for multivariate normal distributions," *Annals of Mathematical Statistics*, 22, 327-350.

[2] Anderson, T. W. and Herman Rubin (1949): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Mathematical Statistics*, 20, 46-63.

[3] Anderson, T. W. and Herman Rubin (1950): "The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochast equations," *Annals of Mathematical Statistics*, 21, 570-582.

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

[5] Hansen, Lars Peter (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029-1054.

[6] Johansen, Søren (1988): "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

[7] Johansen, Søren (1991): "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models," *Econometrica*, 59, 1551-1580.

[8] Johansen, Søren (1995): *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*, Oxford: Oxford University Press.

[9] Magnus, J. R., and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York: John Wiley and Sons.

[10] Sims, Christoper A. (1980): "Macroeconomics and reality," *Econometrica*, 48, 1-48.

[11] Tso, M. K.-S. (1981): "Reduced-rank regression and canonical analysis," *Journal of the Royal Statistical Society, Series B*, 43, 183-189.