Reduced-Rank Regression

Applications in Time Series

Raja Velu

rpvelu@syr.edu

Whitman School of Management

Syracuse University

Gratefully to my teachers GCR & TWA

June 6, 2008

Outline

- Brief Review of Reduced Rank Regression
- VAR Model Structure and RR Features
- Numerical Example
- Recent Developments - Forecasting/ RRR-Shrinkage

Review of RRR Results

- Anderson(1951)

$$Y_t \quad = CX_t + DZ_t + \epsilon_t \qquad (1)$$

$$m \times 1 \quad n_1 \times 1 \quad n_2 \times 1$$

$$Rank\ (C) = r\ < m \qquad (2)$$

- Two implications:

$$l_i^{'}\, C = 0 \quad i = 1, 2, \ldots (m - r) \qquad (3)$$

$$C = AB \qquad (4)$$

Review (contd.)

- Estimation of Anderson's Model

Minimize,

$$|W| = \left|\frac{1}{T}(Y - ABX - DZ)(Y - ABX - DX)'\right|$$

$$= \left|\widetilde{\Sigma}_{\epsilon\epsilon} + (\widetilde{C} - AB)\widehat{\Sigma}_{xx.z}(\widetilde{C} - AB)'\right| \qquad (5)$$

$\widetilde{\Sigma}_{\epsilon\epsilon}-$Residual Covariance matrix under full rank

$\widehat{\Sigma}_{xx.z}-$ Partial Covariance matrix

Thus, SVD on

$\widetilde{\Sigma}_{\epsilon\epsilon}^{-1/2}\ \widetilde{C}\ \widehat{\Sigma}_{xx.z}^{1/2}$ yields Gaussian Estimates

Review (contd.)

Let $\widehat{R}$= $\widetilde{\Sigma}_{\epsilon\epsilon}^{-1/2}\ \widetilde{C}\ \ \widehat{\Sigma}_{xx.z}\ \widetilde{C}'\widetilde{\Sigma}_{\epsilon\epsilon}^{-1/2}$; Let V_j be a normalized eigen vector of $\widehat{R}$ associated with j^{th} largest eigen value, $\widehat{\lambda}_j^2$

$$\widehat{C} = \widehat{A}\widehat{B} = P\widetilde{C} \tag{6}$$

$$\widehat{A} = \widetilde{\Sigma}_{\epsilon\epsilon}^{-1/2}\widehat{V}_{(r)}, \quad \widehat{B} = \widehat{V}'_{(r)}\widetilde{\Sigma}_{\epsilon\epsilon}^{-1/2}\widetilde{C}$$

where $\widehat{V}_{(r)} = \left[\widehat{V}_1, \ldots\ldots\ldots\widehat{V}_r\right]$,

- Estimation of $l_i^{'}\ s$ in (3)

$$\widehat{l}_i' = \widehat{V}'_{(i)}\widetilde{\Sigma}_{\epsilon\epsilon}^{-1/2} \qquad i = r+1, \ldots\ldots\ldots m \tag{7}$$

- Unknown Parameters : $mn_1 \to r(m + n_1 - r)$

Review (contd.)

- Note $\left|\widehat{\Sigma}_{\epsilon\epsilon}\right| = \left|\widetilde{\Sigma}_{\epsilon\epsilon}\right| \prod_{j=r+1}^{m} (1 + \widehat{\lambda}_j^2)$ (8)

$$(1 + \widehat{\lambda}_j^2) = 1/(1 - \widehat{\rho}_j^2)$$

$\widehat{\rho}_j^2$: Squared Partial Canonical Correlations

- LR Test for $H_0 : Rank \ \ (C) \leq r$

$$[T - n_1 + (n_1 - m - 1)/2] \, . \sum_{j=r+1}^{m} \log(1 - \widehat{\rho}_j^2) \tag{9}$$

$$\sim \chi^2_{(m-r)(n_1-r)}$$

- Hence, RRR $\Leftrightarrow$ Canonical Correlation Analysis

VAR Models

- Number of Parameters depend on order '*p*' and dimension '*m*'
- Stationarity imposes constraints on the AR Coefficient matrices
- Anderson (2002) : Canon Corr / VAR   RR
- Box & Tiao (1977): Cannon Corr/Non-stationarity – Cointegration ideas

VAR/RR - Features

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \epsilon_t \tag{10}$$

- Φ_2 – reduced-rank; Φ_1 –Full rank – Anderson's Model
- Rank (Φ_1: Φ_2) lower rank;
- $$Y_t = A(B_1 : B_2) \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \end{pmatrix} + \epsilon_t \tag{11}$$

  (Velu et al 1986)
- $$Y_t = A_1 B Y_{t-1} + A_2 B Y_{t-2} + \epsilon_t \tag{12}$$

  (Reinsel 1985) -Index Model
- Let $\Phi_j = A_j B_j$ and range $(A_{j+1}) \subset range\ (A_j)$

  (Ahn and Reinsel, 1988) –Nested RR

VAR- Unit Root Features

- Model (10) : $Y_t - \Phi_1 Y_{t-1} - \Phi_2 Y_{t-2} = \Phi(L) Y_t = \epsilon_t$
  $|\Phi(L)| = 0$,has $d \leq m$ unit roots; rest are outside unit circle

- Error Correction (Granger) form:
  $$W_t = CY_{t-1} - \Phi_2 W_{t-1} + \epsilon_t \quad (13)$$
  $$Rank\ (C) = Rank\ (-\ \Phi\ (1)) \leq r = m - d \quad (14)$$

- Thus, non-stationarity can be checked via reduced-rank test; Anderson (1951)

- Implications of (14); $W_t = Y_t - Y_{t-1}$ is stationary

VAR- Unit Root Features - Interpretation

- Johansen (1988,1991)

  Note (13) is: $W_t = ABY_{t-1} - \Phi_2 W_{t-1} + \epsilon_t$..........$(14)$

- BY_t are stationary; rows of B are co-integrating vectors
- Let $\underset{m\times 1}{Z_t} = \begin{pmatrix} B \\ B^* \end{pmatrix} Y_t, \quad B^*\ (d \times m)$ is orthogonal to B
- $B^* Y_t$ are non-stationary – “Common Trends”

VAR-Unit Root-RRR - Advantages

- Parsimonious Parametrization
- Exhibit any “co-features” or “common features”
- Improves forecasting
- Other References are :

  Tiao and Tsay (1988)

  Hannan and Deistler (1988)

  Engle and Kozicki (1993)

  Min and Tsay (2005)

Cointegration-Unit Root-Tests

- H_0: $Rank\ (C) = r$ vs $H_1 : Rank\ (C) < r$

- Johansen's Trace Statistics:
$$L_{tr}(r) = -(T - mp) \sum_{j=r+1}^{m} \ln(1 - \widehat{\rho}_j^2) \ldots \ldots\ldots\ldots\ldots (9)$$

- Limiting distribution of L_{tr} is a standard Brownian motion, depends only on 'r'

- Estimation of C, A, B etc. all follow from Anderson (1951)

Beyond Cointegration – Common Features

- Note that in Model
  $W_t = ABY_{t-1} - \Phi_2 W_{t-1} + \epsilon_t,$
  the component BY_{t-1} is stationary
- Find $l_i' \ni l_i'A = 0$ spans the co-feature space (Anderson 1951)
- Find $l_i' \ni l_i'A = 0$ and $l_i'\Phi_2 = 0$, Serial Correlation Common Feature (Engle & Kozicki 1993)
- Common Seasonality (Hecq et al 2006)
- Review Paper : Urga (2007)

Numerical Example

- Four Quarterly Series : Private Investment, GNP, Consumption and Unemployment,1951-1988,
  T= 148 (See Figure 1)
- VAR(2) is specified; also $rank\ (\Phi_2) = 1$ based on partial Canonical Correlations (See Table 2)
- Recall that Partial Canonical analysis between W_t and Y_{t-1}, given $W_{t-1} \Leftrightarrow LR$ for Unit Roots;

  Results suggests $d = 2$ unit roots,thus $r = 2$ cointegrating ranks; (See Table 3)
- Model (14) :$W_t = CY_{t-1} - \Phi_2 W_{t-1} + \epsilon_t$

Figure 1

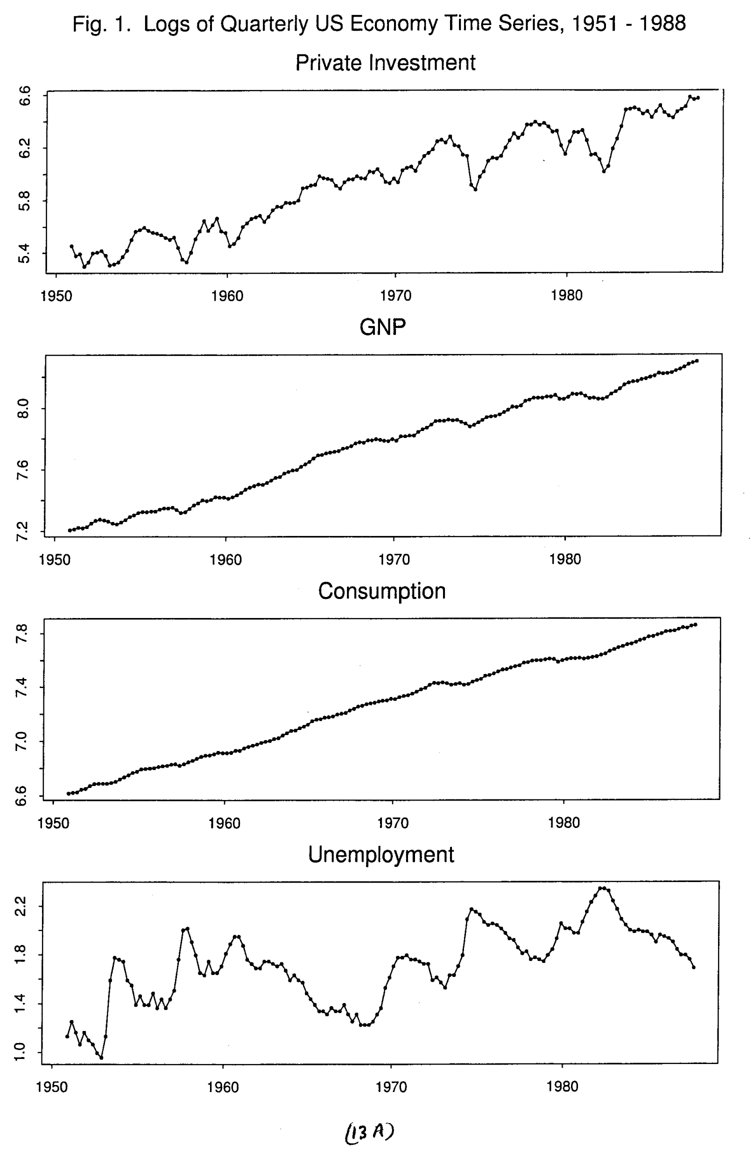


Fig. 1. Logs of Quarterly US Economy Time Series, 1951 - 1988
Private Investment
6.6
6.2
5.8
5.4
1950
1960
1970
1980
GNP
8.0
7.6
7.2
1950
1960
1970
1980
Consumption
7.8
7.4
7.0
6.6
1950
1960
1970
1980
Unemployment
2.2
1.8
1.4
1.0
1950
1960
1970
1980
(13 A)

Table 2

SPECIFICATION OF AR ORDER AND REDUCED RANKS

Table 2. Summary of Test Statistics for AR Order j and Ranks $r_j = m - s$ of Coefficients Φ_j for US Economy Data, Partial Canonical Correlation Analysis between Y_t and Y_{t-j}, given $Y_{t-1}, \ldots$.

j	Eigenvalues	$C(j,s)$			
		$s=4$ $(r=0)$	$s=3$ $(r=1)$	$s=2$ $(r=2)$	$s=1$ $(r=3)$
1	0.999, 0.931, 0.825, 0.540	1895.95	734.00	356.22	109.88
2	0.374, 0.057, 0.004, 0.001	72.69	8.77	0.70	0.09
3	0.098, 0.051, 0.027, 0.002	24.23	10.71	3.88	0.23
4	0.107, 0.071, 0.011, 0.001	25.34	10.99	1.61	0.16
5	0.146, 0.033, 0.012, 0.002	25.02	5.81	1.67	0.22
6	0.102, 0.064, 0.012, 0.004	22.10	9.58	1.85	0.45

- Eigenvalues are squared partial canonical correlations $\hat{\rho}_i^2(j)$. Test for H_0: rank$(\Phi_j) = m - s$ is $C(j,s) = -(T - j - mj - 1) \sum_{i=(m-s)+1}^{m} \log[1 - \hat{\rho}_i^2(j)] \sim \chi^2_{s^2}$

Table 3

SPECIFICATION OF *rank(C)* OR NUMBER *d* OF UNIT ROOTS

Table 3. Summary of LR Tests for rank($\mathbf{C}$) = r (cointegration) or number of unit roots $d = m - r$ for US Economy Data, Partial Canonical Correlation Analysis between W_t and Y_{t-1}, given $W_{t-1}, \ldots$.

j	Eigenvalues	LR statistics r=0 (d=4)	r=1 (d=3)	r=2 (d=2)	r=3 (d=1)
1	0.304, 0.115, 0.060, 0.011	81.20	28.36	10.56	1.56
2	0.207, 0.086, 0.053, 0.012	56.43	22.81	9.72	1.78
3	0.196, 0.085, 0.074, 0.010	56.60	25.24	12.52	1.41
4	0.219, 0.078, 0.072, 0.011	59.20	23.93	12.26	1.59
5	0.179, 0.089, 0.066, 0.013	52.80	24.84	11.57	1.83
6	0.196, 0.099, 0.068, 0.008	56.50	25.75	11.11	1.17

- Eigenvalues are squared partial canonical correlations $\hat{\rho}_i^2$ of W_t and Y_{t-1},

  $-(T-j)\log(\lambda) = -(T-mp)\sum_{i=r+1}^{m}\log(1-\hat{\rho}_i^2)$ is LR statistic.

Numerical Example (contd.)

- Estimation with 'Cointegrating' Reduced Rank Imposed

$$\widehat{C}=\widehat{A}\widehat{B}=\begin{bmatrix} -0.241 & 0 \\ -0.022 & 0 \\ -.01 & .107 \\ -.209 & -.921 \end{bmatrix}\begin{bmatrix} 1 & 0 & -.997 & .092 \\ 0 & 1 & -.942 & .094 \end{bmatrix}$$

$$= \begin{bmatrix} -0.241 & 0 & 0.24 & -.02\,22 \\ -0.022 & 0 & .022 & -.002 \\ -0.01 & 0.107 & -.09 & .009 \\ -0.209 & -0.921 & 1.07 & -0.105 \end{bmatrix}$$

$$-\widehat{\Phi}_2 = \begin{bmatrix} 1 \\ .167 \\ .070 \\ -1.673 \end{bmatrix}\begin{bmatrix} .239 & 0 & 2.724 & 0 \end{bmatrix} = \begin{bmatrix} 0.239 & 0 & 2.724 & 0 \\ .04 & 0 & 0.454\,91 & 0 \\ 0.017 & 0 & 0.190\,68 & 0 \\ -0.400 & 0 & -4.557\,3 & 0 \end{bmatrix}$$

$\left|\widehat{\Sigma}_\epsilon\right| = 4.503 \times 10^{-15}$ -Estimates $\sim$Full Rank LSES

Numerical Example (contd.)

- Implications / Interpretation
- Cointegration : Stationary Combinations:

$$\overset{*}{Z}_{3t} = Y_{1t} - 0.997Y_{3t} + 0.042Y_{4t}$$

$$\overset{*}{Z}_{4t} = Y_{2t} - 0.942Y_{3t} + 0.094Y_{4t}$$

- Common Trend Components:

  $\overset{*}{Z}_{1t} = b_1^{'} Y_t$ and $\overset{*}{Z}_{2t} = b_2^{'} Y_t$

  where $b_1^{'} A$ and $b_2^{'} A = 0$;

  choose $b_1 \ni b_1^{'} \Phi_2 = b_1^{'} a_3 d' = 0$

  Thus, $b_1^{'} W_t = b_1^{'} \xi_t$, so

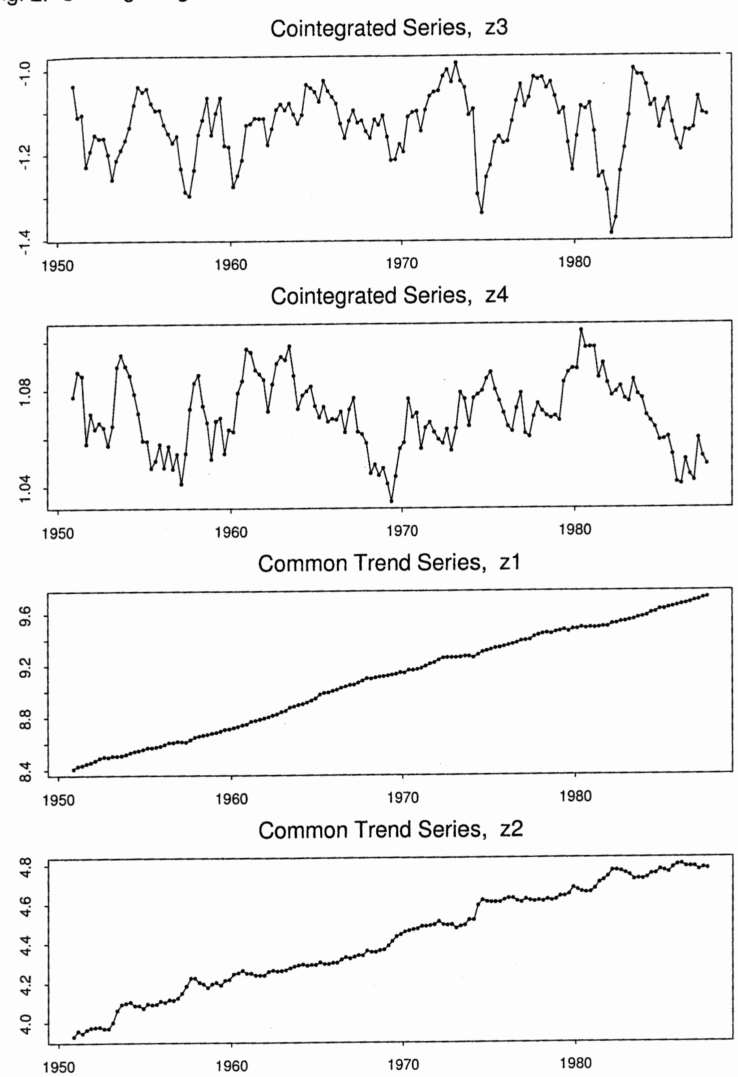
  $\overset{*}{Z}_{1t} = b_1^{'} Y_t$ is a random walk

  $\overset{*}{Z}_{1t} = -0.132Y_{1t} + Y_{2t} + 0.285Y_{3t} + 0.033Y_{4t}$

  See Figure 2

Figure 2

Fig. 2. Cointegrating & Common Trend Series for US Economy Data, 1951 - 1988



Numerical Example (contd.)

- Overall interpretation;

$$\overset{*}{Z}_t = \binom{B}{B^*} Y_t = \binom{Z_{1t}}{Z_{2t}}$$

$$\overset{*}{Z}_{1t} - \overset{*}{Z}_{1t-1} = (B^* a_3) d' W_{t-1} + \epsilon^*_{1t}$$

$$\overset{*}{Z}_{2t} = (I + BA) \overset{*}{Z}_{2t-1} + (B a_3) d' W_{t-1} + \epsilon^*_{2t}$$

$$B^* a_3 = \begin{bmatrix} 0 \\ -.289 \end{bmatrix}, \; B a_3 = \begin{bmatrix} 0.776 \\ -.055 \end{bmatrix},$$

$$I + BA = \begin{bmatrix} .750 & -.191 \\ -.032 & .814 \end{bmatrix}$$

$\overset{*}{Z}_{1t} \sim RW$; $\overset{*}{Z}_{2t} - \overset{*}{Z}_{2t-1} \sim$ with AR(1) structure, $\overset{*}{Z}_{3t} \sim$ AR(2);

$\overset{*}{Z}_{4t} \sim$ AR(1)

Forecasting-Applications

- Effect of RR estimation on prediction, $Rank\ (C) = 1$; $C = \alpha\beta'$

- Covariance matrix of the prediction error:
$$(1 + x_0^{'}(XX^{'})^{-1} x_0)\Sigma_{\epsilon\epsilon} - \left\{ x_0^{'}(XX^{'})^{-1}x_0 - \frac{(\beta^{'} x_0)^2}{\beta^{'}(XX^{'})^{-1}\beta} \right\} (\Sigma_{\epsilon\epsilon} - \Sigma_{\epsilon\epsilon}^{1/2} V_1 V_1^{'} \Sigma_{\epsilon\epsilon}^{1/2})$$
(Reinsel and Velu, 1998, p46)

- Empirical Analysis of Hog data indicate accounting for both non-stationarity and rank reduction consistently outperforms well at all forecasting horizons (Wang and Bessler 2004)

RRR and Shrinkage

- Recall $\widehat{C} = \widehat{A}\widehat{B} = \widehat{\Sigma}_{yy}^{1/2}\widehat{V}_*^{(r)}\widehat{V}_*^{(r)'}\Sigma_{yy}^{-1/2}\widetilde{C}$

  $= \widehat{H}_1^{-1}\widehat{\Delta}_1\widehat{H}_1\widetilde{C}$

  where $\widehat{\Delta}_1 = Diag(1, 1, ..., 1, 0...0), m \times m$ diagonal matrix

  GCV - Breiman and Friedman (1997)

  Reinsel (1998)

$\widehat{\Delta}_1 = Diag(\widehat{d}_j)$, $\widehat{d}_j = \frac{(1-p_1)(\widehat{\rho}_j^2 - p_1)}{(1-p_1)^2\widehat{\rho}_j^2 + p_1^2(1-\widehat{\rho}_j^2)} \qquad j = 1 \ldots m$

where $p_1 = n_1/(T - n_2)$

Forecasting Applications -Contd.

- Forecasting Large Datasets (Carriero, Kapetanoos & Marcellino 2007)

  -RRR (Reinsel and Velu, 1998)

  -BVAR (Doan et al, 1984)

  -BRR (Geweke, 1996)

  -RRP (Restrictions on the posterior mean)

  -MB (Brieman, 1996)

  -FM (Stock and Watson 2002)

  Data: 52 US Macroeconomic series (Real, Monetary &Financial)

- Three Benchmarks: AR (1), BVAR and RW;

- Among the six models: BRR -Short horizons

  RRP -Long Horizons

  FM- Best for 1-step ahead.

- Conclusion: Using Shrinkage and Rank reduction in combination

Future Work

- Shrinkage -RRR, Needs further investigation
  Yuan et al 2007-RRR/ LASSO
- Impact of Rank Misspecification : Anderson (2002)
- More broadly, in VAR context, the impact of misspecification of both the order 'p' and rank 'r' need to be studied.