Reduced Rank Vector Exponential Smoothing

Ashton de Silva

Department of Econometrics and Business Statistics, Monash University

email: Ashton.desilva@buseco.monash.edu.au

May 2004

Abstract

The single source of error state space Vector Simple Exponential Smoothing model is introduced. A hypothesis free two-stage methodology that detects and defines the reduced rank of a smoothing matrix is proposed. A singular value decomposition of the smoothing matrix forms the basis of both stages.

1 Introduction

Exponential Smoothing (ES) has proved to be an effective forecasting tool (Hyndman et al, 2002). The specification however remains essentially univariate with little development of a multiple series approach (Harvey (1986) and Pfeffermann & Allon (1989) are 2 exceptions). I propose a Vector Exponential Smoothing (VES) specification that is a natural extension to the methodology. The VES approach provides the ability to utilise information contained in alternative series whilst maintaining the successful ES structure.

A reduced rank VES model is defined by a singular smoothing coefficient matrix. One advantage of this reduced rank specification is the ability to capture inter-series dependency without increasing the number of unknown parameters exponentially.

A two-stage procedure that determines the rank of the smoothing matrix and defines the nature of dependency is proposed. The singular value decomposition of the smoothing coefficient matrix forms the basis of both stages.

This paper is divided into two sections beginning with the identification of the Vector Simple Exponential Smoothing (VSES) model. This is followed by an outline of the procedure. The procedure naturally divides into two stages. The first stage determines the rank of the smoothing matrix by minimising the AIC. The second stage identifies the collinear structure of the smoothing coefficient matrix for two situations: when the rank of the smoothing coefficient matrix is 1 and when the rank is greater than 1.

2 Vector Simple Exponential Smoothing

The ES methodology was first postulated by Brown (1959). Since this introduction, ES models have proved to be a competitive forecasting tool. Recently in the IJF-M3 competition the ES methodology was very successful (Hyndman et al, 2002) proving to be the most accurate at 6 period ahead forecasts. The state space ES methodology was not ranked below 4 for forecasts ranging from 1 to 8 horizons. However ES has consistently under performed with respect to long-term forecasts. The VES extension of the methodology provides an opportunity to address this apparent shortfall. The multivariate state space form of a simple exponential smoothing model is defined below.

$$\begin{aligned} X_t &= S_{t-1} + E_t,\ E_t \sim^{iid} MVN(0, \Sigma) \\ S_t &= S_{t-1} + \Theta E_t \end{aligned} \tag{1}$$

where

- k denote the number of series;
- X_t denotes the observations of the individual series at time t, dimension $k \times 1$;
- S_t denotes the state vector at time t, dimension $k \times 1$;
- Θ denotes a $k \times k$ matrix of smoothing parameters.

The expanded form of a bivariate Vector Simple Exponential Smoothing (VSES) model is:

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} s_{1t-1} \\ s_{2t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} s_{1t} \\ s_{2t} \end{bmatrix} = \begin{bmatrix} s_{1t-1} \\ s_{2t-1} \end{bmatrix} + \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \tag{3}$$

$$s_{i,t} = s_{i,t-1} + \theta_{i,1} e_{1,t} + \theta_{i,2} e_{2,t} \tag{4}$$

Forecasts are given by

$$\hat{x}_{i,t}(h) = s_{i,t} \tag{5}$$

This model simplifies to k univariate models when Θ and Σ are both diagonal. Forecasts are expected to be at least as accurate as the nested simple univariate ES models.

One particular advantage of the VSES model is its potential to capture a common structure between series. In particular collinear dependency within coefficient matrices have proved particularly useful in long-term forecasting (Engle & Yoo. 1987). The proposed model bears some similarities with the ECM introduced by Engle & Granger (1987). Furthermore the ability to model a collinear structure within a specific state has inferential advantages over other multivariate time series techniques.

2.1 Reduced rank VSES

When Θ has rank $q < k$ it can be decomposed into two $k \times q$ matrices. (Reinsel, 1998; Johansen, 1989):

$$\Theta = \alpha\beta^T \tag{6}$$

Thus β describes the column structure of Θ. Therefore when Θ is singular the state space system (1) can be re-specified as:

$$\begin{aligned} X_t &= HS_{t-1} + E_t,\ E_t \sim^{iid} MVN(0, \Sigma) \\ S_t &= FS_{t-1} + \alpha\beta^T E_t \end{aligned} \tag{7}$$

In the procedure, I fix β and estimate only α. Treating β as fixed decreases the number of parameters to be estimated. The state equation for the bivariate model is now given by:

$$\begin{aligned} s_{1,t} &= s_{1,t-1} + \alpha_{11}(\beta_{11}e_{1,t} + \beta_{12}e_{2,t}) \\ s_{2,t} &= s_{2,t-1} + \alpha_{12}(\beta_{11}e_{1,t} + \beta_{12}e_{2,t}) \end{aligned} \tag{8}$$

An ECM type structure is depicted when β_{11} and β_{12} have opposite signs. By applying a svd of Θ, the rank of Θ and the composition of β can be determined.

$$\begin{aligned} svd(\Theta) &= UDV^T \qquad (9) \\ \Theta &= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{bmatrix} \end{aligned}$$

The svd (9) contains information on the rank and the column structure of Θ. Schott (1997) defines U and V as the orthonormal bases for the column and row space of Θ respectively. The matrix D (a diagonal matrix) contains the same number of nonzero diagonal values as the rank of the matrix. Each value d provides a measure of the level of information pertaining to the associated vectors of U and V. A d-value of zero suggests there is no information. The decomposition is in decreasing order, ie. $d_1 \geq d_2 \geq d_3$.

3 Two-Stage Procedure

3.1 Stage 1: Rank determination

When the rank of Θ is q (where $0 < q < k$) only the first q column/row vectors of V and U are required to specify Θ. In particular the first q columns of V are used to specify and capture the column structure of Θ. To determine the rank of Θ sequential evaluation of the AIC is proposed increasing the number of d-values restricted to be zero each time. The procedure after the initial estimation of (1) and calculation of the AIC is as follows. For $j = k, k-1, \ldots, 1$

1. Set the j^{th} diagonal value (d_j) to zero;
2. Recalculate $\Theta = UDV^T$
3. Recalculate Σ (from equations (1));
4. Calculate the AIC:

$$AIC = log(|\hat{\Sigma}_{EE}|) + \frac{2}{T}qk \tag{10}$$

Compare the AICs, and select q according to the smallest AIC. The AIC comprises 2 parts. The first part Σ_{EE} represents the accuracy of the model whilst $\frac{2}{T}qk$ penalises the model according to the number of estimated parameters. If the rank of the Θ matrix is q, restricting the last $k-q$ diagonal values will not alter $|\Sigma_{EE}|$ but decrease the penalty $\frac{2}{T}qk$. Alternatively if the d_{q-1} diagonal was set to zero then $|\Sigma_{EE}|$ would increase.

3.2 Stage 2: Linear Structure

Once the rank (q) of Θ is chosen the corresponding columns of V are used to form β. The first q vectors of V can be employed directly to form β however this may prove difficult to interpret. The explicit identification of which columns are linearly related and their loading when the linear dependency is characterised as a function of one column only is outlined below. The situation of $q = 1$ is considered first followed by $q > 1$.

3.2.1 q=1

A $k \times k$ matrix with a rank of 1 occurs when $k-1$ of the columns of Θ are proportional to the 1 remaining column. Two examples are presented below for the bivariate VSES model.

Example 1

$$\Theta = \begin{bmatrix} 0.4 & -0.4 \\ 0.4 & -0.4 \end{bmatrix} \tag{11}$$

$$svd(\Theta) = \begin{bmatrix} -0.71 & -0.71 \\ -0.71 & 0.71 \end{bmatrix} \begin{bmatrix} 0.8 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.71 & -0.71 \\ 0.71 & -0.71 \end{bmatrix}^T \tag{12}$$

The linear structure captured by the first column of V is after when it is normalised by setting the first row value to 1. Extracting and normalising the first column of V yields:

$$\beta = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{13}$$

The state equations of the VSES system (1) can then be re-expressed in terms of the linear structure of Θ.

$$\begin{aligned} s_{1,t} &= s_{1,t-1} + \alpha_{11}(e_{1,t} - e_{2,t}) \\ s_{2,t} &= s_{2,t-1} + \alpha_{12}(e_{1,t} - e_{2,t}) \end{aligned} \tag{14}$$

The state system (14) depicts the dependency between the disturbances of both series. The VSES model is now estimated with only 2 unknown smoothing coefficients instead of 4.

Example 2

$$\Theta = \begin{bmatrix} 0.35 & 0.7 \\ 0.15 & 0.3 \end{bmatrix} \tag{15}$$

$$svd(\Theta) = \begin{bmatrix} -0.92 & -0.39 \\ -0.39 & 0.92 \end{bmatrix} \begin{bmatrix} 0.85 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.45 & 0.89 \\ -0.89 & -0.45 \end{bmatrix}^T \tag{16}$$

Extracting and normalising the first column of V yields the following β.

$$\beta = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \tag{17}$$

This results in the following system where α_{11} and α_{12} are the only smoothing coefficients to be estimated.

$$\begin{aligned} s_{1,t} &= s_{1,t-1} + \alpha_{11}(e_{1,t} + 2e_{2,t}) \\ s_{2,t} &= s_{2,t-1} + \alpha_{12}(e_{1,t} + 2e_{2,t}) \end{aligned} \tag{18}$$

For a general $k \times k$ matrix with rank 1 the above procedure of extracting and normalising the first column of V provides a simple and effective way of defining linear dependence in Θ, and in specifying β.

3.2.2 $1 < q < k$

When Θ has rank q the first q column vectors of V are used to specify β. The linear structure of Θ can be determined by the following procedure if the linear dependency can be characterised in terms of one column only. An example of Θ corresponding to a 3×3 smoothing matrix where the third column is double the first column is used to illustrate the procedure.

$$\Theta = \begin{bmatrix} 0.13 & 0.75 & 0.26 \\ 0.69 & 0.17 & 1.38 \\ 0.86 & 0.33 & 1.72 \end{bmatrix} \tag{19}$$

1. Calculate the svd of Θ;

$$svd(\Theta) = \begin{bmatrix} -0.17 & 0.98 & -0.10 \\ -0.61 & -0.18 & -0.77 \\ -0.77 & -0.07 & 0.63 \end{bmatrix} \begin{bmatrix} 2.53 & 0 & 0 \\ 0 & 0.70 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -0.44 & -0.09 & 0.89 \\ -0.19 & 0.98 & 0.00 \\ -0.88 & -0.17 & -0.45 \end{bmatrix}$$

2. Extract any two of the vectors of V that correspond to nonzero values of D. (Selecting the first two vectors is recommended).

$$V_{\cdot,1:2} = \begin{bmatrix} -0.44 & -0.09 \\ -0.19 & 0.98 \\ -0.88 & -0.17 \end{bmatrix} \tag{20}$$

Denote these vectors as $\nu_{\cdot 1}$ and $\nu_{\cdot 2}$ respectively:

$$\nu_{\cdot 1} = \begin{bmatrix} -0.44 \\ -0.19 \\ -0.88 \end{bmatrix}, \nu_{\cdot 2} = \begin{bmatrix} -0.09 \\ 0.98 \\ -0.17 \end{bmatrix}$$

3. Calculate the individual row ratios of the two vectors.

$$\kappa = \frac{\nu_{\cdot 1}}{\nu_{\cdot 2}} \tag{21}$$

$$\kappa = \frac{\nu_{\cdot 1}}{\nu_{\cdot 2}} = \begin{bmatrix} 5.14 \\ -0.19 \\ 5.14 \end{bmatrix} \tag{22}$$

4. By examining the values in κ the linearly dependent columns are identified. It is shown below that *equivalent* numbers indicate linearly dependent columns. When all the values within κ are unique this indicates the matrix has an alternative linear structure from the one currently investigated. (In this case it is recommended that β be defined as the first q columns of V). Examining κ shows that the 1^{st} and 3^{rd} columns of Θ are linearly dependent (ie. the row ratios are identical). Furthermore the value -0.19 is different from 5.14 suggesting the second column of Θ is linearly independent. As numerical inaccuracies and statistical uncertainty exist the detection of which columns are related in general is determined by the κ values that are the closest. (The difference between the values should be approximately zero.)

5. Based on the above examination the two dimensional row space of Θ can be defined. The ordering of the two vectors is inconsequential. One vector will capture the linear dependency between the first and third columns only whilst the remaining vector will denote the linearly unique contribution of the second column. The linear relationship between the first and third columns is more interpretable if one of the values in the original vector ($v_{\cdot 1}$) is normalised ($v^*_{\cdot 1}$). The contribution of the second column should also be normalised.

$$v_{\cdot 1} = \begin{bmatrix} -0.44 \\ -0.19 \\ -0.88 \end{bmatrix}, v^*_{\cdot 1} = \begin{bmatrix} 1.00 \\ 0.43 \\ 2.00 \end{bmatrix} \tag{23}$$

6. Define β to be the rectangular matrix that corresponds to the linear structure that has been identified, namely the third column is double the first and the second column is linearly independent of both columns. It is recommended that β should be orthogonal.

$$\Theta = \begin{bmatrix} 0.13 & 0.75 & 0.26 \\ 0.69 & 0.17 & 1.38 \\ 0.86 & 0.33 & 1.72 \end{bmatrix} = \begin{bmatrix} 0.13 & 0.75 \\ 0.69 & 0.17 \\ 0.86 & 0.33 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix}$$

The model can now be re-estimated using the decomposition of equation (6) holding β fixed. Due to the reduction in the number of parameters to be estimated this is expected to increase the efficiency. Furthermore evidence suggests long-term forecasts may also improve as a result of such a linear structure.

$$\begin{aligned} s_{1,t} &= s_{1,t-1} + \alpha_{11}(e_{1,t} + 2e_{3,t}) + \alpha_{21}e_{2,t} \\ s_{2,t} &= s_{2,t-1} + \alpha_{12}(e_{1,t} + 2e_{3,t}) + \alpha_{22}e_{2,t} \\ s_{3,t} &= s_{3,t-1} + \alpha_{11}(e_{1,t} + 2e_{3,t}) + \alpha_{23}e_{2,t} \end{aligned} \tag{24}$$

3.2.3 Proof

The verification of steps 3 to 6 are considered in terms of a general matrix of dimension $k \times k$. The svd defined as a set of linear equations is the key. The objective is to express the individual values of $v_{\cdot i}$ as functions of Θ and d_i and thus verifying the steps 3 to 6.

The svd of any matrix is equivalent to the eigenvalue decomposition of the matrix's quadratics: $\Theta\Theta^T$ and $\Theta^T\Theta$. The first quadratic $\Theta\Theta^T$ is associated with the eigenvectors of U whilst $\Theta^T\Theta$ is associated with the eigenvectors of V (where U and V are given in equation (9)). The D matrix contains the positive square roots of the eigenvalues of either quadratic (eigenvalues of $\Theta^T\Theta$ and $\Theta\Theta^T$ are equivalent). By presenting the eigenvalue decomposition as a set of linear equations individual expressions for each loading $v_{\cdot i}$ can be derived.

$$\Theta^T\Theta v_{\cdot i} = d_i v_{\cdot i} \tag{25}$$

The key to this proof is the special structure of $\Theta^T\Theta$. Let Θ denote a $k \times k$ matrix characterised by the p^{th} column being proportional to the j^{th} column by a factor of ρ. The verification relies on the structure in the quadratic $\Theta^T\Theta$. It is shown that the linear structure of $\Theta^T\Theta$ reflects the column structure of Θ.

The structure of $\Theta^T\Theta$ is easily verified by re-parameterising Θ in terms of it columns.

$$\Theta = [\theta_{\cdot 1}\ \theta_{\cdot 2} \ldots \theta_{\cdot j} \ldots \rho\theta_{\cdot j} \ldots \theta_{\cdot k}] \tag{26}$$

$$\Theta^T\Theta = \begin{bmatrix} \theta_{\cdot 1}^T \\ \theta_{\cdot 2}^T \\ \ldots \\ \theta_{\cdot j}^T \\ \ldots \\ \rho\theta_{\cdot j}^T \\ \ldots \\ \theta_{\cdot k}^T \end{bmatrix} [\theta_{\cdot 1}\ \theta_{\cdot 2} \ldots \theta_{\cdot j} \ldots \rho\theta_{\cdot j} \ldots \theta_{\cdot k}] \tag{27}$$

Combining the result derived in equation (27) with equation (25) results in ($\tilde{\theta}_i = \theta_{\cdot i}^T\theta_{\cdot i}$):

$$\begin{bmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \\ \vdots \\ \tilde{\theta}_j \\ \vdots \\ \rho\tilde{\theta}_j \\ \vdots \\ \tilde{\theta}_k \end{bmatrix} v_{\cdot i} = d_i v_{\cdot i} \tag{28}$$

The j^{th} and p^{th} equations of the above systems are:

$$\tilde{\theta}_j v_{\cdot i} = d_i v_{ji} \tag{29}$$

$$\rho\tilde{\theta}_j v_{\cdot i} = d_i v_{pi} \tag{30}$$

Rearranging equation (29) in terms of v_{ji} yields:

$$v_{ji} = \frac{\tilde{\theta}_j v_{\cdot i}}{d_i} \tag{31}$$

Substituting equation (31) into equation (30) (solving for the right hand side of equation (30)) proves the proportional relationship in Θ is mirrored in the values of $v_{\cdot i}$ (for $d_i \neq 0$)

$$v_{pi} = \rho v_{ji} \tag{32}$$

Therefore calculating the elements of κ (equation (33)) reveals the vectors that are linear dependent on one another.

$$\kappa = \frac{v_{p1}}{v_{p2}} = \frac{\rho v_{j1}}{\rho v_{j2}} = \frac{v_{j1}}{v_{j2}} \tag{33}$$

This derivation proves that linear structures can be detected and determined through the svd of Θ. Furthermore it may be possible to parameterise the relationship through the calculation and inference of κ (equations (21) & (33). The loading can then in turn be defined by normalising (step 6 of the procedure) κ with respect to the dependent columns (ie. rearranging equation 32).

$$\rho = \frac{v_{pi}}{v_{ji}} \tag{34}$$

3.3 Implementation Summary

1. Estimate the VSES model.
2. Calculate the AIC.
3. Calculate the svd of Θ.
4. **Stage 1**
   (a) Recalculate Θ from the svd after setting the smallest diagonal (d_k)of D to zero.
   (b) Calculate the AIC from the re-estimated Σ matrix.
   (c) Repeat steps (a) and (b) for d_i, where $i = k-1, \ldots 1$.
   (d) Select q according to the smallest AIC.
5. **Stage 2**
   (a) Extract the first 2 column vectors of V.
   (b) Calculate κ and search for *equivalent* values.
   (c) Define β according to relationships in κ.
6. Re-estimate the re-specified VSES model.

In practice it may be appropriate to skip stage 2 if the rank of Θ is 1 or the underlying linear structure is not deemed to be important.

4 Conclusion

The preceding analysis provides insight into the specification and diagnosing of reduced rank VSES models. The VES methodology is presented in its simplest form. VES and its reduced rank forms offer a viable alternative to other multivariate forecasting approaches. The ability to model common component behaviour without complicated routines makes this approach very attractive.

Research detailing the various forms of the VES methodology is needed and will establish it as a credible multivariate forecasting tool. Furthermore the detection of linear dependency characterised by more than one column also needs to be investigated.

References

[1] Anderson, T. W. (1984). *An introduction to multivariate statistical analysis.* New York, John Wiley & Sons.

[2] Brown, R. G. (1959). *Statistical forecasting for inventory control.* New York, McGraw-Hill.

[3] Engle, R. F., and C. W. J. Granger (1987). "Co-integration and error correction: representation, estimation, and testing." *Econometrica* 55(2): 251-276.

[4] Engle, R. F., and B. S. Yoo (1987). "Forecasting and testing in cointegrated systems." *Journal of Econometrics* 35: 143-159.

[5] Hamilton, J. D. (1994). *Times series analysis*, Princeton University Press.

[6] Harvey, A. (1986). "Analysis and generalisation of a multivariate exponential smoothing model." *Management Science* 32(3): 374-380.

[7] Harvey, A., and S.J. Koopman, Multivariate structural times series models. In *System Dynamics in Economic and Financial Models*, ed. C. Heij, et al. 1997, Chichester: John Wiley & Sons.

[8] Hyndman, R.J., A. B. Koehler, R.D. Snyder and S. Grose (2002). "A state space framework for automatic forecasting using exponential smoothing methods." *International Journal of Forecasting*

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

[10] Lutkepohl,H. (1991). *Introduction to multiple time series Analysis.* Berlin-Heidelberg, Springer-Verlag.

[11] Pfeffermann, D., and J. Allon (1989). "Multivariate exponential smoothing:method and practice." *International Journal of Forecasting* 5: 83-98.

[12] Reinsel, G. C., and R. P. Velu (1998). *Multivariate reduced rank regression.* New York, Springer-Verlag.

[13] Schott, J. (1997). *Matrix analysis for statistics.* New York, John Wiley & Sons.