

# Brief Description of the Work

Jyotishka Ray Choudhury & Aytijhya Saha

## \* Estimation of $W$ in the model $Y = WY + \epsilon$

Let the vector  $X = (X_1, X_2, \dots, X_k)$  and  $\text{cov}(X) = V$  which can be partitioned as:

$$\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$$

where  $V_{11} = \text{cov}(Y_1)$ ,  $V_{22} = \text{cov}(Y_2)$  and  $V_{12} = V_{21} = \text{cov}(Y_1, Y_2)$ , with  $Y_1 = (X_1, X_2)$ ,  $Y_2 = (X_3, \dots, X_k)$ . Then  $V_{11.2} = V_{11} - V_{12}V_{22}^{-1}V_{21}$ . Consider the individual elements of

$$V_{11.2} = \begin{bmatrix} v_{11.2} & v_{12.2} \\ v_{21.2} & v_{22.2} \end{bmatrix}$$

We would like to compare, between  $v_{12.2}/v_{11.2}$  and  $v_{12.2}/v_{22.2}$  and keep whichever is bigger among them and discard the other one. This would give regression coefficient between each pair of random variables eliminating the effect of the other variables. This way one can select the coefficients  $W$  matrix which would be non-zero, in the model

$$Y = WY + \epsilon$$

and would be a way to see directional dependence in the fixed time period. Then we would move the window. This would give a way to do path analysis.

This process can be compared with other ones, such backward substitution or forward selection, eliminating the multicollinearity effect.

## Implementation

For each district, we started with only those districts which we found out to be possible regressors for that district according to the comparison rule stated above, and performed backward regression. We have performed the computation of the estimated  $W$  matrix for 14 overlapping timespans, viz.

- Day 41 to Day 100
- Day 71 to Day 130
- Day 100 to Day 160

- $\vdots$
- $\vdots$
- Day 400 to Day 460
- Day 430 to Day 490

To view the regression matrices along with colour codes arranged based on distance-induced suitable ordering, [please click here](#).

Each colour code corresponding to the intersection of a specific row and a specific column represents a different measure of geographical distance between the corresponding districts. The codes can be found in the following diagram:

0	1	2	3	4	5	6	7	8	9

Figure 1: Colour codes

Note that if there are (nearly)  $k$  districts in-between two specific districts, the corresponding entry is coloured by the  $k$ -th colour from the above list.

## \* Estimation of the original spatio-temporal model $Y_t = WY_t + A_1Y_{t-1} + A_2Y_{t-2} + \varepsilon_t$

Method of estimation of  $W$  is discussed in section 1.

We have been working with the following spatio-temporal model:

$$Y_t = WY_t + A_1Y_{t-1} + A_2Y_{t-2} + \varepsilon_t$$

where  $\varepsilon_t$  is the unknown error vector at time  $t$ . Note that in this model,  $W$ ,  $A_1$ , and  $A_2$  are unknown, time-invariant,  $d \times d$  matrices, i.e. parameters of the spatio-temporal model.

As of now, we shall assume that the error vectors over time are independently and identically distributed. We can perform the following calculation:

$$(I - W)Y_t = A_1Y_{t-1} + A_2Y_{t-2} + \varepsilon_t \implies Y_t = B_1Y_{t-1} + B_2Y_{t-2} + \delta_t$$

where  $B_1$ ,  $B_2$ ,  $\delta_t$  are obtained by pre-multiplying  $A_1$ ,  $A_2$ ,  $\varepsilon_t$  by  $(I - W)^{-1}$ . The equation on the LHS is the actual model, and the transformed one on the RHS is a reduced model.

Note that the reduced model is nothing but an usual VAR(2) model, with parameters  $B_1$  and  $B_2$ . So, we can simply estimate  $B_1$  and  $B_2$  using traditional methodologies. Let us denote their estimates by  $\hat{B}_1$  and  $\hat{B}_2$  respectively.

Now, we can estimate the error vectors in the following manner:

$$(I - W)^{-1}\varepsilon_t = \delta_t = Y_t - (\hat{B}_1 Y_{t-1} + \hat{B}_2 Y_{t-2})$$

We shall be making an assumption that the covariance matrix of  $\varepsilon_t$ , say  $V$ , is a diagonal matrix.

Taking covariance on both sides, we get that:

$$\begin{aligned} (I - W)^{-1}\varepsilon_t &= Y_t - (\hat{B}_1 Y_{t-1} + \hat{B}_2 Y_{t-2}) \\ \implies \text{cov}((I - W)^{-1}\varepsilon_t) &= \text{cov}(Y_t - (\hat{B}_1 Y_{t-1} + \hat{B}_2 Y_{t-2})) \\ \implies (I - W)^{-1}V_\varepsilon((I - W)^{-1})^T &= \text{ResCov}_Y \\ \implies (I - W)^T V_\varepsilon^{-1} (I - W) &= \text{ResCov}_Y^{-1} \end{aligned}$$

We shall be writing the last equation as:  $M^T V M = C$ , where we are defining:  $M := I - W$ ,  $V := V_\varepsilon^{-1}$ , and  $C := \text{ResCov}_Y^{-1}$ .

This is to be noted that by our assumption, the following structures have been imposed on the matrices that we are supposed to estimate:

- The principal diagonal of  $W$  should contain only 0's. Naturally, the principal diagonal of  $M$  should contain only 1's.
- $V_\varepsilon$  is, by assumption, a diagonal matrix with strictly positive entries. Thus, the matrix  $V$  should be a diagonal matrix, with its diagonal entries equal to the reciprocals of the diagonal entries of  $V_\varepsilon$ .
- Being a covariance matrix,  $\text{ResCov}_Y$  is a known, symmetric, positive definite matrix (That is to be estimated from data). Since the inverse of a positive definite matrix is necessarily positive definite,  $C$  is a known, symmetric, positive definite matrix.
- All the matrices mentioned above are  $d \times d$  matrices.  $M$  and  $V$  are to be estimated, and  $C$  is a known, symmetric, positive definite matrix.

### Estimation for two regimes

We can partition our time series dataset into multiple regimes, and estimate  $M$  and  $V$  accordingly. For simplicity, we may consider two regimes. In general, for the  $s$ -th regime, we shall be denoting the corresponding matrices by  $M$ ,  $V_s$  and  $C_s$ . Note that  $M$  is independent of  $s$ , because we are assuming that the spatial matrix  $W$  remains invariant over time.

In the case of  $s = 2$ , i.e. two regimes, we are required to solve for  $M$ ,  $V_1$  and  $V_2$  the following system of equations:

$$M^T V_1 M = C_1 \quad \text{and} \quad M^T V_2 M = C_2 \quad (\clubsuit)$$

where the structures of all the matrices have been discussed earlier.

The task of estimation can be performed in the following way:



Figure 2: Time series plots 1

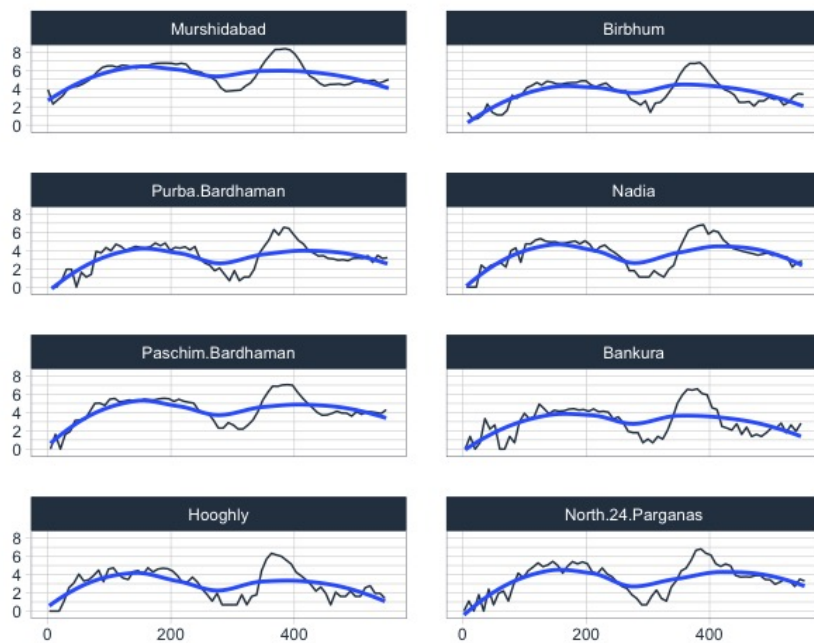


Figure 3: Time series plots 2

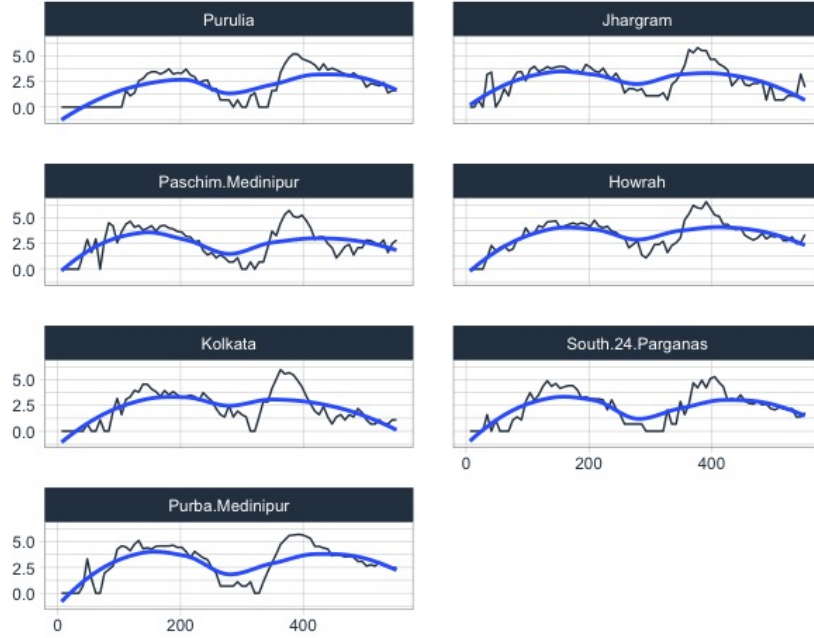


Figure 4: Time series plots 3

- We can start off by writing down a spectral decomposition of  $C_1$ , let's say  $C_1 = D_1^T \Sigma_1 D_1$  where  $\Sigma_1$  is a diagonal matrix containing the eigenvalues of  $C_1$ , say  $\lambda_1, \lambda_2, \dots, \lambda_d$ , and the rows of  $D_1$  contain the corresponding eigenvectors.
- Define  $V_1 := \text{diag}(\lambda_1^2, \lambda_2^2, \dots, \lambda_d^2)$ , viz. a diagonal matrix containing the squared eigenvalues of  $C_1$  as its principal diagonal elements.
- Define  $M := V_1^{-\frac{1}{2}} \Sigma_1^{\frac{1}{2}} D_1$ . Note that:

$$M^T V_1 M = D_1^T \Sigma_1^{\frac{1}{2}} V_1^{-\frac{1}{2}} V_1 V_1^{-\frac{1}{2}} \Sigma_1^{\frac{1}{2}} D_1 = D_1^T \Sigma_1 D_1 = C_1$$

- Define  $V_2 := (M^{-1})^T C_2 M^{-1}$ .

This is to be noted that our chosen  $V_1$ ,  $V_2$  and  $M$  satisfy all the desired properties mentioned earlier. So, our chosen  $W$ ,  $V_1$  and  $V_2$  are solutions to the system of equations (\*).

### \section{\textbf{\* Prediction

We shifted the data by 1, and performed a log-transformation of the data.

VAR(2) model was fitted with the data of first 400 days, and predicted next 154 days using the model.

We calculated prediction errors of each district using the formula:

$$\frac{1}{154} \sum_{i=441}^{554} \left| \frac{\text{predicted}_i - \text{actual}_i}{\text{actual}_i} \right|$$