EM estimation of the Dynamic Coregionalization Model with varying coefficients[1]

Francesco Finazzi and Alessandro Fassò
University of Bergamo - Dept. IIMM, alessandro.fasso@unibg.it

Abstract: The satellites from NASA's Earth Science Project Division, like AURA, produce data for the concentration of various airborne pollutants. Calibrating satellite data using ground level monitoring networks and other meteorological and land characterizing variables is mandatory. To do this, it is important to use an approach which is able to manage large datasets coming from different sources, structural missingness and spatial and temporal correlation. In this paper, we extend the Dynamic Coregionalization Model introduced in Fassò and Finazzi (2011) to the case of space-time varying coefficients in order to increase the model flexibility and to make it suitable for large regions such as Europe.

Keywords: air quality monitoring, missing data, dynamic coregionalization

1 Introduction

The Dynamic Coregionalization Model (DCM) of Fassò and Finazzi (2011) has been proven to be quite appropriate for modeling multivariate space-time environmental data in the non-collocated case and in the presence of missing data. When data are collected over continent-size regions, the statistical model considered must be enough flexible to accommodate for local conditions. In order to gain this flexibility, the DCM is extended here to the case of varying coefficients. The model is described in Section 2 and its estimation is addressed in Section 3.

2 The varying coefficients model

Let $\mathbf{y}(\mathbf{s},t) = (y_1(\mathbf{s},t),...,y_q(\mathbf{s},t))$ be the $q-$variate response variable at site $\mathbf{s} \in \mathcal{D} \subset \mathcal{R}^2$ and time $t \in N^+$. The model equation is

$$\mathbf{y}(\mathbf{s},t) = \mathbf{X}(\mathbf{s},t) \cdot \left[\mathbf{K_x}\beta + \mathbf{K_z}\mathbf{z}(t) + \sum_{j=1}^{c} \gamma_j \mathbf{K}_{\mathbf{w}}^{j} \mathbf{w}^{j}(\mathbf{s},t) \right] + \varepsilon(\mathbf{s},t) \qquad (1)$$

[1]This research is part of Project EN17, 'Methods for the integration of different renewable energy sources and impact monitoring with satellite data', Lombardy Region under 'Frame Agreement 2009'

where $\mathbf{X}(\mathbf{s},t)$ is a matrix of known coefficients (for instance $\mathbf{X}(\mathbf{s},t) = \mathbf{I}_q \otimes \mathbf{x}(\mathbf{s},t)$ is the $q \times (bq)$ diagonal block matrix built from the $1 \times b$ covariate vector $\mathbf{x}(\mathbf{s},t)$), $\mathbf{z}(t)$ is a latent $p-$dimensional temporal state with markovian dynamics $\mathbf{z}(t) = \mathbf{G}\mathbf{z}(t-1) + \eta(t)$ with $\mathbf{G}$ a stable transition matrix and $\eta \sim N(0, \mathbf{\Sigma}_\eta)$ while each $\mathbf{w}^j(\mathbf{s},t) = \left(w_1^j(\mathbf{s},t), ..., w_q^j(\mathbf{s},t)\right)$, $1 \leq j \leq c$ is a q-dimensional gaussian latent coregionalization component with covariance and cross-covariance matrix function $\mathbf{\Gamma}_j = cov\left(w_i^j(\mathbf{s},t), w_{i'}^j(\mathbf{s}',t)\right) = \mathbf{V}_j \rho_j(h, \theta_j)$, $1 \leq i, i' \leq q$, $1 \leq j \leq c$. Each $\mathbf{V}_j$ is a correlation matrix and each ρ_j is a valid correlation function parametrized by θ_j. Finally, $\varepsilon(\mathbf{s},t) = (\varepsilon_1(\mathbf{s},t), ..., \varepsilon_q(\mathbf{s},t))$ is the measurement error which is assumed to be white-noise in space and time with $\varepsilon_i(\mathbf{s},t) \sim N(0, \sigma^2_{\varepsilon,i})$, $1 \leq i \leq q$.

The matrices $\mathbf{K}_\mathbf{x}$, $\mathbf{K}_\mathbf{z}$ and $\mathbf{K}_\mathbf{w}^j$ are matrices of known coefficients which guarantee conformability of the model equation (1) and acts as selection matrices with respect to the columns of $\mathbf{X}(\mathbf{s},t)$. The model parameter set is $\Psi = \{\beta, \sigma^2_\varepsilon, \mathbf{G}, \mathbf{\Sigma}_\eta, \gamma, \mathbf{V}, \theta\}$ where $\beta = (\beta_1, ..., \beta_q)'$, $\sigma^2_\varepsilon = \left\{\sigma^2_{\varepsilon,1}, ..., \sigma^2_{\varepsilon,q}\right\}$, $\gamma = \{\gamma_j, ..., \gamma_c\}$, $\theta = \{\theta_1, ..., \theta_c\}$ and $\mathbf{V} = \{\mathbf{V}_1, ..., \mathbf{V}_c\}$.

3 Likelihood function and missing data

At each time t, each variable y_i is observed over the set of spatial sites $\mathcal{S}_i = \{\mathbf{s}_{i,1}, ..., \mathbf{s}_{i,n_i}\}$, $1 \leq i \leq q$. The sets in $\mathcal{S} = \{\mathcal{S}_1, ..., \mathcal{S}_q\}$ are not constrained and can be disjoint. The observed vector at time t is then $\mathbf{y}_t(\mathcal{S}) = (\mathbf{y}_{1,t}(\mathcal{S}_1), ..., \mathbf{y}_{q,t}(\mathcal{S}_1))'$ $= \mathbf{y}_t$ and it has dimension $N = n_1 + ... + n_q$. The observation equation is $\mathbf{y}_t = \mu_t + \varepsilon_t$, where $\mu_t = \mathbf{U}_{\mathbf{x},t}\beta + \mathbf{U}_{\mathbf{z},t}\mathbf{z}_t + \gamma_1 \mathbf{U}^1_{\mathbf{w},t}\mathbf{w}^1_t + ... + \gamma_c \mathbf{U}^c_{\mathbf{w},t}\mathbf{w}^c_t$, $\mathbf{U}_{\mathbf{x},t} = \mathbf{X}_t \mathbf{K}_\mathbf{x}$, $\mathbf{U}_{\mathbf{z},t} = \mathbf{X}_t \mathbf{K}_\mathbf{z}$ and $\mathbf{U}^j_{\mathbf{w},t} = \mathbf{X}_t \mathbf{K}^j_\mathbf{w}$.

In the definition of the likelihood function, the distributions involved are

$$\begin{aligned}
\left(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{w}^1_t, ..., \mathbf{w}^c_t\right) &\sim N_N\left(\mu_t, \mathbf{\Sigma}_\varepsilon\right) \\
\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) &\sim N_p\left(\mathbf{G}\mathbf{z}_{t-1}, \mathbf{\Sigma}_\eta\right) \\
\mathbf{w}^j_t &\sim N_N\left(0, \mathbf{\Sigma}^j\right), \ 1 \leq j \leq c
\end{aligned}$$

Let $\mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_T)$, $\mathbf{Z} = (\mathbf{z}_1, ..., \mathbf{z}_T)$ and $\mathbf{W}^j = \left(\mathbf{w}^j_1, ... \mathbf{w}^j_T\right)$. The complete-data log-likelihood function is given by:

$$-2l\left(\Psi; \mathbf{Y}, \mathbf{Z}, \mathbf{W}^1, ..., \mathbf{W}^c\right) = T \log|\mathbf{\Sigma}_\varepsilon| + \sum_{t=1}^T (\mathbf{y}_t - \mu_t)' \mathbf{\Sigma}_\varepsilon^{-1} (\mathbf{y}_t - \mu_t) +$$

$$T \log|\mathbf{\Sigma}_\eta| + \sum_{t=1}^T (\mathbf{z}_t - \mathbf{G}\mathbf{z}_{t-1})' \mathbf{\Sigma}_\eta^{-1} (\mathbf{z}_t - \mathbf{G}\mathbf{z}_{t-1}) + \sum_{j=1}^c T \log\left|\mathbf{\Sigma}^j\right| \sum_{t=1}^T \left(\mathbf{w}^j_t\right)' \left(\mathbf{\Sigma}^j\right)^{-1} \mathbf{w}^j_t$$

At each time t, the observation vector $\mathbf{y}_t$ can be partitioned in the following way: $\mathbf{y}^*_t = \left[\begin{array}{cc} \mathbf{y}^{(1)}_t & \mathbf{y}^{(2)}_t \end{array}\right]'$ where $\mathbf{y}^{(1)}_t = \mathbf{L}_t \mathbf{y}_t$ is the sub-vector of the non-missing data and $\mathbf{L}_t$ is the selection matrix of the observed data at time t. The vector $\mathbf{y}^*_t$ is a permutation of $\mathbf{y}_t$ and $\mathbf{y}_t = \mathbf{D}_t \cdot \begin{bmatrix} \mathbf{y}_t^{(1)} & \mathbf{y}_t^{(2)} \end{bmatrix}'$, where $\mathbf{D}_t$ is a permutation matrix. The partitioned measurement equation becomes $\mathbf{y}_t^{(l)} = \mu_t^{(l)} + \varepsilon_t^{(l)}$, $l = 1, 2.$ and the variance-covariance matrix of the permuted errors is conformably partitioned, namely $Var\left[\left(\varepsilon_t^{(1)}, \varepsilon_t^{(2)}\right)'\right] = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix}$. In what follows, $\mathbf{Y}^{(1)} = \left(\mathbf{y}_1^{(1)}, ..., \mathbf{y}_T^{(1)}\right)$ is the collection of the observed data.

4 EM estimation

At the E-step of the EM algorithm, the following conditional expectation is evaluated:

$$
\begin{aligned}
Q\left(\Psi, \Psi^{(k)}\right) &= E_{\Psi^{(k)}}\left[-2l\left(\Psi; \mathbf{Y}, \mathbf{Z}, \mathbf{W}^1, ..., \mathbf{W}^c\right) \mid \mathbf{Y}^{(1)}\right] \\
&= E_{\Psi^{(k)}}\left[E_{\Psi^{(k)}}\left[-2l\left(\Psi; \mathbf{Y}, \mathbf{Z}, \mathbf{W}^1, ..., \mathbf{W}^c\right) \mid \mathbf{Y}^{(1)}, \mathbf{Z}, \mathbf{W}^1, ..., \mathbf{W}^c\right] \mid \mathbf{Y}^{(1)}\right] \\
&= T \log |\mathbf{\Sigma}_\varepsilon| + tr\left(\mathbf{\Sigma}_\varepsilon^{-1} \sum_{t=1}^{T} \mathbf{\Omega}_t\right) + \\
&\quad T \log |\mathbf{\Sigma}_\eta| + tr\left\{\mathbf{\Sigma}_\eta^{-1}\left(\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{G}' - \mathbf{G}\mathbf{S}_{10}' + \mathbf{G}\mathbf{S}_{00}\mathbf{G}'\right)\right\} + \\
&\quad \sum_{j=1}^{c} T \log \left|\mathbf{\Sigma}^j\right| \cdot tr\left\{\left(\mathbf{\Sigma}^j\right)^{-1} \sum_{t=1}^{T} \mathbf{w}_t^{j,T} \cdot \left(\mathbf{w}_t^{j,T}\right)' + \mathbf{A}_t^{j,T}\right\}
\end{aligned}
$$

where:

$$
\begin{aligned}
\mathbf{\Omega}_t &= E_{\Psi^{(k)}}\left[\mathbf{e}_t \cdot \mathbf{e}_t' + \mathbf{\Lambda}_t \mid \mathbf{Y}^{(1)}\right] = E_{\Psi^{(k)}}\left[\mathbf{e}_t \cdot \mathbf{e}_t' \mid \mathbf{Y}^{(1)}\right] \\
&= \mathbf{D}_t \begin{bmatrix} \mathbf{\Omega}_t^{(11)} & \mathbf{\Omega}_t^{(11)}\mathbf{R}_{11}^{-1}\mathbf{R}_{21} \\ \mathbf{R}_{21}\mathbf{R}_{11}^{-1} & \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{\Omega}_t^{(11)}\mathbf{R}_{11}^{-1}\mathbf{R}_{21} + \left(\mathbf{R}_{22} - \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\right) \end{bmatrix} \mathbf{D}_t' \\
\mathbf{e}_t &= E_{\Psi^{(k)}}\left[\mathbf{y}_t - \mu_t \mid \mathbf{Y}^{(1)}, \mathbf{Z}, \mathbf{W}^1, ..., \mathbf{W}^c\right] \\
&= \mathbf{D}_t \begin{bmatrix} \mathbf{y}_t^{(1)} - \mu_t^{(1)} \\ \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\left(\mathbf{y}_t^{(1)} - \mu_t^{(1)}\right) \end{bmatrix} \\
\mathbf{\Lambda}_t &= Var_{\Psi^{(k)}}\left[\mathbf{y}_t - \mu_t \mid \mathbf{Y}^{(1)}, \mathbf{Z}, \mathbf{W}^1, ..., \mathbf{W}^c\right] \\
&= \mathbf{D}_t \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{22} - \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12} \end{bmatrix} \mathbf{D}_t' \\
\mathbf{\Omega}_t^{(11)} &= E_{\Psi^{(k)}}\left[\mathbf{e}_t^{(1)} \mid \mathbf{Y}^{(1)}\right] \cdot E_{\Psi^{(k)}}\left[\mathbf{e}_t^{(1)} \mid \mathbf{Y}^{(1)}\right]' + Var_{\Psi^{(k)}}\left[\mathbf{e}_t^{(1)} \mid \mathbf{Y}^{(1)}\right] \\
\mathbf{w}_t^{j,T} &= E_{\Psi^{(k)}}\left(\mathbf{w}_t^j \mid \mathbf{Y}^{(1)}\right); 1 \leq j \leq c \\
\mathbf{A}_t^{j,T} &= Var_{\Psi^{(k)}}\left(\mathbf{w}_t^j \mid \mathbf{Y}^{(1)}\right); 1 \leq j \leq c
\end{aligned}
$$

Moreover, $\mathbf{z}_t^T = E_{\Psi^{(k)}}\left(\mathbf{z}_t \mid \mathbf{Y}^{(1)}\right)$ and $\mathbf{P}_t^T = Var_{\Psi^{(k)}}\left(\mathbf{z}_t \mid \mathbf{Y}^{(1)}\right)$ are given by the

Kalman smoother output and

$$\mathbf{S}_{11}=\sum_{t=1}^{T}\mathbf{z}_t^T\left(\mathbf{z}_t^T\right)'+\mathbf{P}_t^T;\ \mathbf{S}_{10}=\sum_{t=1}^{T}\mathbf{z}_t^T\left(\mathbf{z}_{t-1}^T\right)'+\mathbf{P}_{t,t-1}^T;\ \mathbf{S}_{00}=\sum_{t=1}^{T}\mathbf{z}_{t-1}^T\left(\mathbf{z}_{t-1}^T\right)'+\mathbf{P}_{t-1}^T$$

The maximization step of the EM algorithm involves the minimization

$$\Psi^{(k+1)}=\arg\min_{\Psi}Q\left(\Psi,\Psi^{(k)}\right)$$

The estimates $\hat{\theta}^{(k+1)}=\left\{\hat{\theta}^1,...,\hat{\theta}^c\right\}^{(k+1)}$ and $\hat{\mathbf{V}}=\left\{\hat{\mathbf{V}}^1,...,\hat{\mathbf{V}}^c\right\}^{(k+1)}$ are obtained by numerical minimization. The close form solutions for $\hat{\mathbf{G}}^{(k+1)}$ and $\hat{\mathbf{\Sigma}}_\eta^{(k+1)}$ are already given in Fassò and Finazzi (2011) while the solution for the remaining parameters are obtained by solving $\frac{\partial Q\left(\Psi,\Psi^{(k)}\right)}{\partial\Psi}=0$ and they are

$$\begin{aligned}
\left(\hat{\sigma}_{i,\varepsilon}^2\right)^{(k+1)} &= \frac{tr\left(\sum_{t=1}^{T}\mathbf{\Omega}_t|_{i,i}\right)}{Tn_i}\\
\hat{\beta}^{(k+1)} &= \left[\sum_{t=1}^{T}\left(\mathbf{U}_{\mathbf{x},t}'\mathbf{U}_{\mathbf{x},t}\right)\right]^{-1}\cdot\left[\sum_{t=1}^{T}\mathbf{X}_{\mathbf{x},t}'\left(\mathbf{e}_t^T+\mathbf{U}_{\mathbf{x},t}'\beta^{(k)}\right)\right]\\
\hat{\gamma}_i^{(k+1)} &= \frac{tr\left[\sum_{t=1}^{T}\left(\mathbf{F}_t^T-\mathbf{G}_t^T-\mathbf{H}_t^T\right)\right]}{tr\left[\sum_{t=1}^{T}\mathbf{U}_{\mathbf{w},t}\left(\mathbf{w}_t^{i,T}\cdot\left(\mathbf{w}_t^{1,T}\right)'+\mathbf{A}_t^{i,T}\right)\mathbf{U}_{\mathbf{w},t}'\right]}
\end{aligned}$$

for each $1\le i\le q$, with $\mathbf{\Omega}_t|_{i,i}$ the $i-th$ diagonal block of $\mathbf{\Omega}_t$. Moreover

$$\begin{aligned}
\mathbf{F}_t^T &= \left(\mathbf{e}_t^T+\gamma_i\mathbf{U}_{\mathbf{w},t}^i\mathbf{w}_t^{i,T}\right)\left(\mathbf{w}_t^{i,T}\right)'\left(\mathbf{U}_{\mathbf{w},t}^i\right)'\\
\mathbf{G}_t^T &= 2\sum_{j\neq i}^{c}\gamma_j\mathbf{U}_{\mathbf{w},t}^iCov_{\Psi^{(k)}}\left(\mathbf{w}_t^i,\mathbf{w}_t^j\mid\mathbf{Y}^{(1)}\right)\left(\mathbf{U}_{\mathbf{w},t}^j\right)' \quad (2)\\
\mathbf{H}_t^T &= 2\mathbf{U}_{\mathbf{z},t}Cov_{\Psi^{(k)}}\left(\mathbf{z}_t,\mathbf{w}_t^i\mid\mathbf{Y}^{(1)}\right)\left(\mathbf{U}_{\mathbf{w},t}^i\right)' \quad (3)
\end{aligned}$$

and the conditional covariances in (2) and (3) are computed straightforwardly from the multivariate Gaussian distribution of the joint $(\mathbf{y}_t,\mathbf{w}_t,\mathbf{z}_t)$.

References

Fassò A. and Finazzi F. (2011) Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data. Environmetrics. In printing.