

Technical Results.

Youngsoo Baek.

Jul 26th, 2018.

This is a file to collate theoretical/technical results that arise on the way.

Heuristic support for not using binary adjacency matrix in Griffith (2013)

In his paper (2013), Griffith proposes employing similar scheme as Moran's basis expansion in a T -point time series setting. The equivalent of spatial adjacency matrix \mathbf{A} specified in this case is derived from the well-known Durbin-Watson statistic. That is, the $T \times T$ matrix \mathbf{C}_T is specified as follows:

$$\mathbf{C}_T = \begin{bmatrix} 1 & -1 & 0 & \dots & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & \dots & 0 \\ 0 & -1 & 2 & -1 & 0 & \dots & 0 \\ \vdots & & & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & -1 & 2 & -1 \\ 0 & \dots & \dots & \dots & \dots & -1 & 1 \end{bmatrix}.$$

In this way, $\mathbf{x}'(\mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}')'\mathbf{C}_T(\mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}')\mathbf{x}$ yields the Durbin-Watson statistic, where $\mathbf{x} = (x_1, \dots, x_t)'$ is the column vector containing observations at time t .

Now, Griffith recites Basilevsky (1983) to state that \mathbf{C}_T has analytically known eigenvectors: that is, the j -th eigenvector for \mathbf{C}_T is given by

$$\mathbf{e}_j = \frac{1}{\sqrt{T+1}} \left(\sin \frac{1}{T+1} j\pi, \dots, \sin \frac{T}{T+1} j\pi \right)'.$$

Then, he says the following:

"The number of positive and of negative eigenvalues equals $(T-1)/2$ if T is odd; when T is even, the number of positive eigenvalues is $(T-2)/2$ and the number of negative eigenvalues is $(T-1)/2$."

I am yet uncertain whether this remark has foundations in Basilevsky (1983). This is confusing to me, since \mathbf{C}_T **should only have nonnegative eigenvalues, and, furthermore, the Moran basis expansion matrix also should only have nonnegative eigenvalues.** I outline a sketch of the relatively straightforward proof below.

Proposition 1. \mathbf{C}_T is positive semidefinite.

Proof. Let a T -by- T matrix \mathbf{C}_T be defined as in Griffith (2013). It can be readily shown that this matrix is a Laplacian matrix of a linear, undirected simple graph $G = (V, E)$, where the set of T vertices consists of each time period $t = 1, \dots, T$. The binary adjacency matrix \mathbf{A} of this graph can be specified as follows:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \dots & \dots & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & 1 & 0 & 1 \\ 0 & \dots & \dots & \dots & \dots & 1 & 0 \end{bmatrix}.$$

Since \mathbf{C}_T is a Laplacian matrix, it is also positively semidefinite.

As a corollary, this implies that **all eigenvalues of \mathbf{C}_T are nonnegative**. \square

We now proceed to show that the Moran's basis expansion on \mathbf{C}_T also yields only non-negative eigenvalues.

Proposition 2. Let \mathbf{X} be a full-rank, $T \times p$ real matrix of covariates, collected over T discrete periods. Let \mathbf{P}_x be the projection matrix of \mathbf{X} . Then, a T -by- T matrix defined by:

$$\mathbf{M} \equiv (\mathbf{I} - \mathbf{P}_x)' \mathbf{C}_T (\mathbf{I} - \mathbf{P}_x),$$

is also positively semidefinite.

Proof. It is trivial to show that \mathbf{M} is a real symmetric matrix. Since it is a Hermitian matrix, and it is known that both $\mathbf{I} - \mathbf{P}_x$ and \mathbf{C}_T are positively semidefinite, we can apply the well-known results and conclude that \mathbf{M} is positively semidefinite. \square

R simulations seem to support the above statements; in particular, Griffith's own example of Puerto Rico, which consists of 168 time points, yielded only positive eigenvalues through eigendecomposition functions in R.

If my above results are true, this provides some theoretical support as to why Griffith's recommended \mathbf{C}_T is not appropriate in the broader context of spatiotemporal modelling. In his paper, the spatiotemporal adjacency structure is specified as a Kronecker product of terms involving \mathbf{C}_T and the areal adjacency matrix \mathbf{A} . However, given the usually high dimensions there is a need to choose proper number of eigenvectors r to include in the model. Hughes and Haran (2013) suggested a reasonable criterion of selecting eigenvectors corresponding to positive eigenvalues, since it is usually expected that neighboring areas (both in sense of time and space) should behave alike. Griffith's version of \mathbf{C}_T , however, is ill-suited for this purpose in time context, since all eigenvalues are positive. Though it is uncertain how it affects the spatiotemporal modelling, there does not seem to be a good rationale for preferring it over other forms of adjacency matrices that account for inherent seasonal structures of the time series.

Extending the latent spatial state model in MSTM suggested by Bradley et al. (2016) to order p .

Bradley et al. (2016) provides a detailed specification of the Multivariate Spatio-Temporal Model (MSTM) for gaussian process. Here, the process model for \mathbf{y}_t can be thought of in a state-space model framework as follows:

$$\text{Output model: } \mathbf{y}_t = \mathbf{X}_t\beta_t + \mathbf{S}_{X,t}\eta_t + \xi_t.$$

$$\text{State model: } \eta_t = \mathbf{M}_{B,t}\eta_{t-1} + \mathbf{u}_t.$$

Here, according to their model notation, η_t is the spatial random effects vector of length $r \ll n_t$, where n_t is the number of spatial locations where process is observed at time t . The state model representation is possible in this form since η_t is modeled by a VAR(1) time series model with propagator matrix $\mathbf{M}_{B,t}$. The authors assure that η_{t-1} and \mathbf{u}_t for $t = 2, \dots, T$ are uncorrelated by the construction of the propagator matrix, which capitalizes on desirable orthogonality properties of a Moran's basis expansion.

Some of the economic variables in the QWI data they studied in the paper have more complex seasonal/temporal dependence structure than that of AR order 1, which motivates the need to extend the model for η_t into VAR model of order $p > 1$. Specifically, suppose $p = 2$. Using the same notation, we now need to specify two propagator matrices:

$$\text{VAR}(2) \text{ specification: } \eta_t = \mathbf{M}_t\eta_{t-1} + \mathbf{N}_t\eta_{t-2} + \mathbf{u}_t.$$

It is no longer straightforward to construct t -variant \mathbf{M} and \mathbf{N} as the authors proposed in the paper; as for one example, we can now allow the two propagator matrices to be *both orthogonally restricted to $C(\mathbf{B}_t)$, but not necessarily to each other*. This intuitively is understandable since given the dependence across far-reaching temporal lags for η_t , the variability explained by each of those components will be reasonably collinear with each other.

State representation of VAR(p) model.

Suppose we have properly specified M_t and N_t . One of the desirable properties of these propagator matrices is that their residing in the orthogonal subspace of $C(\mathbf{B}_t)$. This really means that their construction is \mathbf{X}_t -dependent.

Now, the VAR(2) model for η_t can be represented in a state-space model form (Brockwell and Davis, 2016) as follows:

$$\begin{bmatrix} \eta_t \\ \eta_{t-1} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_t & \mathbf{N}_t \\ I_r & O_r \end{bmatrix} \begin{bmatrix} \eta_{t-1} \\ \eta_{t-2} \end{bmatrix} + \begin{bmatrix} \xi_t \\ \mathbf{0} \end{bmatrix}, \quad t = T-1, \dots, 1,$$

where I_r is an $r \times r$ identity matrix, and O_r is an $r \times r$ matrix of zeros.

Redefining the notation:

$$H_t \equiv \Phi_t H_{t-1} + \Xi_t, \quad t = T-1, \dots, 1.$$

In this state-model representation, we can now apply the well-known Kalman smoother techniques in the MCMC scheme when fitting the model. Forward filtering-backward sampling is a MCMC technique that can be used to sample H_t in the Gibbs sampler (Carter and Kohn, 1994). Details about the FFBS scheme specific to this model are also given in Bradley et al. (2016).

To apply the Kalman smoother, we need knowledge about the covariance structures of H_t . In particular, denote $\text{Var}(H_t)$ by Σ_t , and $\text{Cov}(H_t, H_{t-1})$ by $\Gamma(t, t-1)$. These matrices have a block structure relating to the covariance matrices of η_t :

$$\Sigma_t = \begin{bmatrix} \text{Var}(\eta_t) & \text{Cov}(\eta_t, \eta_{t-1}) \\ \text{Cov}(\eta_t, \eta_{t-1}) & \text{Var}(\eta_{t-1}) \end{bmatrix},$$

$$\Gamma(t, t-1) = \begin{bmatrix} \text{Cov}(\eta_t, \eta_{t-1}) & \text{Cov}(\eta_t, \eta_{t-2}) \\ \text{Var}(\eta_{t-1}) & \text{Cov}(\eta_{t-1}, \eta_{t-2}) \end{bmatrix}.$$

Since $\eta_t \sim \text{Gaussian}(\mathbf{0}, \mathbf{P}_t)$, assuming η_1 and η_2 are uncorrelated, we have the following relations:

$$\begin{aligned} \text{Cov}(\eta_t, \eta_{t-1}) &= \mathbf{M}_t \mathbf{P}_t + \mathbf{N}_t \text{Cov}(\eta_{t-1}, \eta_{t-2}), \quad \text{Cov}(\eta_3, \eta_2) = \mathbf{M}_3 \mathbf{P}_2, \\ \text{Cov}(\eta_t, \eta_{t-2}) &= \mathbf{M}_t \text{Cov}(\eta_{t-1}, \eta_{t-2}) + \mathbf{N}_t \mathbf{P}_{t-2}, \quad \text{Cov}(\eta_3, \eta_1) = \mathbf{N}_3 \mathbf{P}_1. \end{aligned}$$

If $t = 1$; that is, if \mathbf{X} is invariant over time, the above relations can be simplified since the propagator matrices \mathbf{M} and \mathbf{N} are also invariant over time. This can be the case when the exogenous variables included in the model are simply an intercept term or dummy variables for intervention control. We can furthermore then define \mathbf{P} as also time-invariant: $\mathbf{P} \equiv \sigma_K^2 \mathbf{P}^*$, where $\mathbf{P}^* \equiv (\mathbf{S}'\mathbf{Q}\mathbf{S})^{-1}$, as in original model specification.

$$\begin{aligned} \text{Cov}(\eta_t, \eta_{t-1}) &= \sum_{i=0}^{t-3} \mathbf{N}^i \mathbf{M} \mathbf{P}, \\ \text{Cov}(\eta_t, \eta_{t-2}) &= \sum_{i=0}^{t-3} \mathbf{M} \mathbf{N}^i \mathbf{M} \mathbf{P} + \mathbf{N} \mathbf{P}, \end{aligned}$$

where $\mathbf{N}^0 \equiv \mathbf{I}_r$.

Specifying propagator matrices.

In this state-space representation, we can now apply the derivation process of the authors of the MSTM paper for latent spatial process modeled by VAR(2). Again, as the authors acknowledge, this derivation process has no apparent spatial interpretation and purely facilitates the specification of propagator matrices that are often not straightforward from the given data. The set-up assumptions of the modeling, e.g. specifying exogenous variables to be t -invariant, or assuring the dimension of the bases $r \ll n$, will assure that the computations needed in this process are feasible.

In a state-space representation, we can rewrite the model formula as below:

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_t & \mathbf{O}_p \\ \mathbf{O}_n & \mathbf{X}_{t-1} \end{bmatrix} \begin{bmatrix} \beta_t \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{S}_{X,t} & \mathbf{O}_r \\ \mathbf{O}_n & \mathbf{S}_{X,t-1} \end{bmatrix} \left(\begin{bmatrix} \mathbf{M}_t & \mathbf{N}_t \\ I_r & O \end{bmatrix} \begin{bmatrix} \eta_t \\ \eta_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_t \\ \mathbf{0} \end{bmatrix} \right) + \begin{bmatrix} \xi_t \\ \xi_{t-1} \end{bmatrix}.$$

Assuming \mathbf{X} is scaled, and defining

$$\Phi_t \equiv \begin{bmatrix} \mathbf{M}_t & \mathbf{N}_t \\ I_r & O \end{bmatrix}$$

and

$$\Theta_t \equiv \begin{bmatrix} \mathbf{S}_{X,t} & \mathbf{O}_r \\ \mathbf{O}_n & \mathbf{S}_{X,t-1} \end{bmatrix},$$

it can be shown that the following holds true:

$$\Theta'_t \tilde{\mathbf{y}}_t = \begin{bmatrix} \Theta'_t \tilde{\mathbf{X}}_t & I_{2r} \end{bmatrix} \begin{bmatrix} \tilde{\beta}_t \\ \tilde{\mathbf{u}}_t \end{bmatrix} + \Phi_t \tilde{\eta}_t,$$

where:

$$\tilde{\mathbf{y}}_t = \begin{bmatrix} \mathbf{y}_t - \xi_t \\ \mathbf{y}_{t-1} - \xi_{t-1} \end{bmatrix}, \tilde{\mathbf{X}}_t = \begin{bmatrix} \mathbf{X}_t \\ \mathbf{X}_{t-1} \end{bmatrix},$$

$$\tilde{\beta}_t = \begin{bmatrix} \beta_t \\ \beta_{t-1} \end{bmatrix}, \tilde{\eta}_t = \begin{bmatrix} \eta_t \\ \eta_{t-1} \end{bmatrix}, \quad \tilde{\mathbf{u}}_t = \begin{bmatrix} \mathbf{u}_t \\ \mathbf{0} \end{bmatrix}.$$

Thus, the propagator matrix Φ_t can be now specified in the similar manner as described in Bradley et al., 2016, using the Moran's operator that expands on the adjacency matrix I_{2r} on the orthogonal subspace of the column space of matrix $\mathbf{B}_t \equiv \begin{bmatrix} \Theta'_t \tilde{\mathbf{X}}_t & I_{2r} \end{bmatrix}$.

Cited Works.

- Basilevsky, A. (1983). *Applied Matrix Algebra in the Statistical Sciences*. North-Holland, New York.
- Bradley, J. R., Holan, S. H., and Wikle, C. K. (2015). Multivariate spatio-temporal models for high- dimensional areal data with application to longitudinal employer-household dynamics. *The Annals of Applied Statistics*, 9(4):1761-1791.
- Brockwell, P. and Davis, R. (2016). *Introduction to Time Series and Forecasting*. Springer, New York.
- Carter, C. K. and Kohn, R. (1994). On Gibbs Sampling for State Space Models. *Biometrika*, 81(3):541-553.
- Griffith, D. A. (2012). Space, time, and space-time eigenvector filter specifications that account for autocorrelation. *Estadstica Espanola*, 54(177):211-237.