Model-based now-casting for surveillance data: large-scale application

Krzysztof Sakrejda[⋆] and Nicholas G. Reich[⋆]

[⋆]University of Massachusetts - Amherst

March 24, 2016

Abstract

Motivation:

1. VAR(p) form a basic class of models used for prediction in multiple contexts (econometric tradition, ecology).
2. After fitting, VAR(p) models are often analyzed for stability (econometric, population biology cites) via a variety of Eigen-decomposition based numerical methods which produce information on long-run growth (stationarity/nonstationarity/near-non-stationarity), oscillations (imaginary eigenvalues), transient dynamics and dampening or instability).
3. An alternative view of these diagnostic numbers sees them as meaningful parameters in a generative model.
4. This alternative view suggests that we could place weak priors on these parameters and encourage models that produce stable numbers.
5. Such models could still be checked for fit and predictive performance
6. In public health such models could be made to match the expert expectations that under certain conditions specific diseases either fade or grow explosively (unless mediated by interventions).
7. Models tuned based on these priors should be viewed as: - descriptive and predictive, encoding an expert-driven bias. - in live prediction exercises, comparing biased model predictions to data is particularly critical to indicate when a model fails to capture unusual features (those that go against regularizing priors).

Methods:

1. VAR(p) models can be re-written as VAR(1) models with a structured projection matrix of coefficients, identity sub-matrices, and zero sub-matrtices.

2. Eigenvalue decomposition of the fitted matrix can be used diagnostically to indicate whether the fitted model represents a stationary system, how fast the stationary system will return to equilibrium after a disturbance, how strong the tendency to oscillate is, what periodicity/ies of oscillations is/are important. These features are used diagnostically after fitting a VAR(p) model to data.

3. Knowledge of infectious disease dynamics can also be encoded using the metrics derived from the components of an Eigenvalue decomposition. For example infectious disease systems are known to be at least weak stationary in endemic settings and nearly non-stationary models *can* be used to model hyperendemic settings.

4. We translate a series of [specific statements] on endemic and hyperendemic infectious disease time series behavior into priors on components of the Eigenvalue decomposition of a projection matrix.

5. We characterize outputs of prior simulations based on reconstructing the projection matrix from these prior components.

6. We fit VAR(p) models to simulated time series including seasonal, stationary, and nearly non-stationary series.

7. We fit VAR(p) models to dengue data, we incorporate a reporting delay model, we compare predictions to SPAMD/prior-unconstrained SARIMA/frequentist SARIMA/MOA

1 Introduction

2 Methods

2.1 VAR(p), VAR(1) formulations

A vector auto-regressive model for K states with p lags, typically written as VAR(p), can be written in one form as:

$$\vec{y}_t = \vec{\mu} + \mathbf{A_1}\vec{y}_{t-1} + \ldots + \mathbf{A_p}\vec{y}_{t-p} + \vec{\epsilon_t} \tag{1}$$

It can also always be re-written as a VAR(1) model after some index-shuffling:

$$\mathbf{Y_t} = \begin{bmatrix} \vec{y}_t \\ \vec{y}_{t-1} \\ \cdots \\ \vec{y}_{t-p+1} \end{bmatrix}, \mathbf{U} = \begin{bmatrix} \vec{\mu} \\ 0 \\ \cdots \\ 0 \end{bmatrix}, \mathbf{Z}_t = \begin{bmatrix} \vec{\epsilon_t} \\ 0 \\ \cdots \\ 0 \end{bmatrix} \tag{2}$$

The VAR(p) model can then be written as a VAR(1) model for Y_t

$$\mathbf{Y}_t = \mathbf{U} + \left[\mathbf{A_1}\mathbf{A_2}\ldots\mathbf{A_p}\right]\mathbf{Y}_{t-1} + \mathbf{Z}_t \tag{3}$$

This variation is interesting because Eigenvalue decomposition can be applied to $\mathbf{A}$ to produce model diagnostic components. The Eigenvalue decomposition separates $\mathbf{A}$ into two matrices. Matrix $\mathbf{B}$ is the same dimensions as $\mathbf{A}$, and its columns are orthonormal and referred to as the Eigenvectors of $\mathbf{A}$. The matrix $\mathbf{\Lambda}$ is diagonal and contains the eigenvalues of $\mathbf{A}$. The decomposition is:

$$\mathbf{A} = \mathbf{B}\mathbf{\Lambda}\mathbf{B}^{-1} \tag{4}$$

From our perspective we could begin with priors on $\mathbf{B}$ and $\mathbf{\Lambda}$, and construct $\mathbf{A}$ in the model.

2.2 Eigenvalue desiderata

We are not looking for a generic prior over all of $\mathbf{B}$ and $\mathbf{\Lambda}$, but for priors on the components and sometimes priors that establish relationships among the components. Some examples:

1. The largest absolute Eigenvalue should be less than 1. That is, the system should return towards a specific mean and finite variance after disturbance.

2. The ratio between the largest absolute Eigenvalue and the next largest should be ”large enough”. That is the return to the stationary distribution should be on a time-scale relevant to the system of interets. Making this statement scale-free in some way (or related to a time-series length/seasonality parameter would be useful).

3. This should be true of all pairs of Eigenvalues—so some prior relationship on all $(K \times p) - 1$ ratios would be useful (for $K \times p$ Eigenvalues).

4. For stable (well, stationary) systems, we are often interested in modeling the occasional excursions from what is ”typical” (i.e.-the typical epidemic is something a healh system is made to handle, the atypical epidemic is what they would like to have predictions for.

5. The dominant right (?) Eigenvector should (for appropriately scaled timeseries used in the model) contain values mostly on the same order of magnitude, although some may be very small. This is an attempt to be precise about the idea that, after scaling, we expect the model to represent mostly time-series which will in the future be non-zero.

6. In population biology there's some theory where the dominant left (?) Eigenvector represents the ”future reproductive value of the current population in each class”. I'm not comfortable with the interpretation in that context so I'll have to nail it down some more...

2.3 Decomposition problems

The *Eigenvalue desiderata* section is written as though we want priors on all of $\mathbf{A}$ in the VAR(1) formulation. However, much of the VAR(1) formulation is book-keeping s.t. the portions of $\mathbf{Y_{t-1}}$ which appear in $\mathbf{Y_t}$ are placed in the the proper entries of $\mathbf{Y_t}$. Dealing with the Eigenvalue decomposition for the full $K * p$ Eigenvalues and Eigenvectors would require us to restrict some values/vectors s.t. the entries of $\mathbf{A}$ dealing purely with indexing were correctly re-created.

In practice we only want to deal with the upper K rows of $\mathbf{A}$ and leave the lower $(K \times p) - K$ rows as specified (sub-diagonal identity sub-matrices and remaining all-zeros). In matrix notation:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \dots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ I_K & 0 & \dots & 0 & 0 & 0 \\ 0 & I_K & 0 & \vdots & \vdots & \vdots \\ \vdots & 0 & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & & & 0 & \vdots \\ 0 & 0 & \dots & 0 & I_K & 0 \end{bmatrix} \tag{5}$$

Which means we only need a decomposition for the $K \times (K \times p)$ matrix $\mathbf{A}^*$, written in matrix notation:

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \ldots & \mathbf{A}_{p-1} & \mathbf{A}_p \end{bmatrix} \tag{6}$$

The Eigenvalue decomposition (EVD) is non-sensical here since the matrix is not square, but the Singular Value Decompositoin (SVD) defines the corresponding decomposition for non-square matrices and should be usable.

- I don't know if the conditions on Eigenvalues/Eigenvectors translate directly into conditions on singular values and vectors from the SVD.
- How many singular values should there be here anyway, K, I think, not $K \times p$, but not sure. This should come out of the SVD definition here.
- Are some of the entries of the singular value vectors going to be zero? Are some singular vectors going to be all-zero? This should also be clear from the SVD definition.

3 Results

4 Discussion