Connectivity Analysis through Structural Vector Auto-Regressive (SVAR) Modeling

G. Chen[1], Z. S. Saad[1], J. P. Hamilton[2], I. H. Gotlib[2], and R. W. Cox[1]
[1]SSCC/DIRP/NIMH, National Institutes of Health, Bethesda, Maryland, United States, [2]Mood and Anxiety Disorders Laboratory, Department of Psychology, Stanford University, Stanford, California, United States

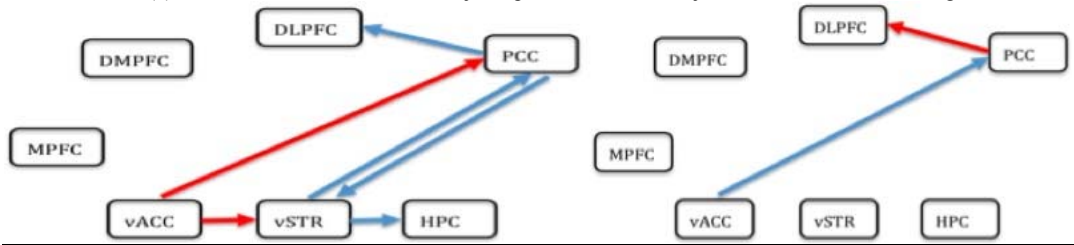
Introduction: Connectivity analysis of FMRI data explores the interactions of regions of interest (ROIs) under various perceptual, cognitive, and affective conditions. Vector autoregression (VAR) and structural equation modeling (SEM) are two popular tools for modeling, at the neural-network level, functional neuroimaging data. As a *data-driven* approach, VAR assumes that effectively connected regions exert *delayed* influences on one another. In contrast, the *hypothesis-driven* SEM is used to validate a connectivity model where connected regions have *contemporaneous* interactions among them. Each approach can identify some features of a network, but they each fail to model other characteristics. For example, the instantaneous correlation among the regions in VAR is deemed of no interest and is relegated to the residual terms. Conversely, SEM assumes that all neural interactions are instantaneous and no lagged correlation is considered within and across regions, an assumption apparently violated with FMRI data of TR ~ 2s. Because lagged effects may dominate the interregional interactions, the covariance structure estimated from the data for SEM might not reflect the true interregional associations, leading to a failure to validate a network and to a biased estimation of path coefficients.

Method: A realistic model of network relations should include both contemporaneous and lagged interactions. Unifying the VAR system and the SEM equations, we present a generalized structural vector auto-regressive (SVAR) model of order p for the time series vector $X(t)$ of n regions, $X(t) = A_0X(t)+A_1X(t\text{-}1)+\ldots+A_pX(t\text{-}p)+\mathbf{c}_1z_1(t)+\ \ldots+\mathbf{c}_qz_q\ (t)+B\boldsymbol{\varepsilon}(t)$, where $n\times n$ matrix A_i codes the effects of history with a lag of i on the current state of the regions (i = 1, …, p), $\mathbf{c}_j$ is the effect vector of size $n\times1$ for the j-th covariate (j = 1, …, q), and $z_1(t)$, …, $z_q\,(t)$, are observable exogenous variables, and usually are determined outside of the network mechanism. $\boldsymbol{\varepsilon}(t) \sim N(0, I_{n\times n})$. B = diag(b_1, b_2, …, b_n). A_0 represents contemporaneous effects as in SEM, while A_1, …, A_p are lagged effects as in VAR(p).

The incorporation of exogenous variables into the SVAR(p) model not only allows minimum data pre-processing (slice timing and head motion correction are sufficient) with no need for band-pass filtering or removal of confounds (*e.g.*, physiological data), but also provides a more statistically robust model. In addition, such an model affords it the capability to address complexities inherent in analyzing FMRI data. For example, if censoring time points (*e.g.*, spikes) is desirable, one can create, for each censored time point, a covariate with the same length as the time series composed of a 1 at the censored time point and zero at all other time points. Another important feature is that, if there are multiple segments of time series data at each ROI from various blocks, runs, or sessions, these segments can be concatenated with corresponding gains in statistical power. To account for discontinuities in the data, we can add some dummy variables at each time discontinuity as covariates in the SVAR(p) model.

We have implemented the above modeling approach in a program, *1dSVAR*, in the open source language R [1] using the *vars* package [2]. The solution of the SVAR(p) equations is achieved through either a scoring algorithm for the maximum likelihood estimates or an optimization scheme based on Nelder-Mead, quasi-Newton and conjugate-gradient algorithms [2]. This implementation allows us not only to combine two modeling methodologies, VAR and SEM, and two analysis strategies, data-driven and model validation, into one framework, but also to estimate both historical and immediate effects in one consistent system. Compared to SEM alone, this should improve power in detecting instantaneous effects among the regions. Compared to VAR alone, this should improve power in estimating lagged effects by modeling instantaneous effects as opposed to relegating them to the errors.

Application: *1dSVAR* was applied to resting-state data from 14 control subjects (18 axial slices, voxel size = $3.44 \times 3.44 \times 5$ mm^3, TR = 1.2 s, TE = 30 ms, flip angle = 77°, FOV = 220 mm, scanning time = 5 min.). ROIs were selected based on the comparison of a seed-based VAR(1) model on whole brain with 16 major depressive disorder subjects. Results are shown in Fig. 1.



To / From	vACC	vSTR	HPC	MPFC	DLPFC	DMPFC	PCC	vACC	vSTR	HPC	MPFC	DLPFC	DMPFC	PCC
vACC	0.046	**0.077**	0.010	-0.036	-0.007	0.028	**0.054**	-	0.047	-	-	-	-	**-0.060**
vSTR	-0.003	**0.168**	**-0.138**	-0.053	-0.089	-0.005	**-0.085**	-	-	0.055	-	-	-	-0.058
HPC	-0.009	-0.049	**0.262**	-0.018	-0.027	0.010	-0.029	-	-	-	-	-	-	-
MPFC	-0.019	-0.017	0.025	**0.400**	-0.015	-0.008	-0.016	-	-	-	-	-	-	-
DLPFC	0.021	-0.010	0.019	-0.004	**0.574**	0.043	-0.060	-	-	-	-	-	-	-
DMPFC	0.020	-0.045	-0.021	0.020	-0.034	**0.435**	-0.005	-	-	-	-	-	-	-
PCC	0.004	**-0.043**	-0.027	-0.020	**-0.154**	-0.025	**0.565**	-	-	-	-	**0.064**	-	-

Fig. 1. The table shows the one-lag path (left, shaded) and instantaneous (right) coefficients estimated with *1dSVAR*. Bold numbers indicate a significance level of two-tailed $p < 0.05$, uncorrected. The SVAR(1) model indicates five significant directional paths: vACC → vSTR, vACC → PCC, vSTR → HPC, vSTR → PCC, and PCC → DLPFC. The network above each effect type is shown with only interregional paths: red, excitatory; blue, inhibitory. For comparison, SEM failed to validate the instantaneous network.

Conclusion: *1dSVAR* in the open source language R has the potential for various uses, for example, the mixture of a data-driven approach for detecting lagged effects (A_i, i = 1, …, p) and a model validation approach for finding instantaneous effects (A_0), a data-driven approach for estimating lagged effects combined with model comparison or model search for finding instantaneous effects, model validation for estimating both instantaneous and lagged effects, model comparison and model validation for both instantaneous and lagged effects, *etc*. We believe that the SVAR modeling strategy holds great promise for both simultaneous modeling of contemporaneous and lagged effects in neural networks.

References

[1] R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org (2010).
[2] B. Pfaff, VAR, SVAR and SVEC Models: Implementation within R package vars. Journal of Statistical Software 27(4)(2008).