<u>Project Summary</u>: Spatial-Econometric Models for the Political & Social Sciences

Spatial interdependence is ubiquitous and central substantively and theoretically across social science. Empirically, clustering of outcomes on some dimension(s), *spatial association*, is also obvious. However, outcomes may exhibit spatial association for three reasons. Units may respond similarly to similar exposure to similar exogenous internal or external stimuli (*common exposure*), units' responses may depend on others' responses (*interdependence*, or *contagion*), or the putative outcome may affect the variable along which clustering manifests (*selection*). We may find states' adoptions of some economic treaty, e.g., to cluster geographically or along other dimensions of *proximity*, e.g., bilateral trade-volume, because proximate states experience similar exogenous domestic or foreign conditions or because each state's decision to sign depends on whether proximate others sign or because signing the treaty spurs bilateral trade. The theories and policy implications that these alternative sources of spatial association support differ starkly.

In previously funded research (Franzese & Hays 2003, 2004ab, 2005ab, 2006abc, 2007abcd, 2008abcd, Franzese et al. 2008), we explored specification, estimation, interpretation, and presentation of spatial and spatiotemporal-lag linear-regression models, which reflect contagion and common exposure directly and so can distinguish them as alternative substantive sources of observed spatial association. We gauged analytically and by simulation the biases of omitting contagion (non-spatial) or of including spatial lags to reflect it but ignoring the lags' simultaneity (naïve spatial), finding the former a typically far graver concern. The latter becomes so too as interdependence strengthens. We explained likelihood and moment estimators that redress the simultaneity, and our simulations showed them near-dominant. We explored various model-specification tests and the sensitivity of estimator performance to misspecification of non-spatial components, of patterns of spatial-connectivity, or of the strict assumptions of the consistent estimators. We then showed how to calculate spatial/spatiotemporal dynamics, multipliers, and effects, along with certainty estimates.

We propose next to undertake a like set of tasks for two more-challenging classes of models: interdependence among qualitative or systems of dependent variables (*S-QualDep* or *S-SysEq*) models, and those that estimate jointly the connectivity pattern and the strength of contagion by that pattern (*Estimated*-**W**), including the case where connectivity is endogenous to the dependent variable (*Endogenous*-**W**) models. As before, for a core target audience of applied researchers in political/social science and political/social methodologists less familiar with spatial/network analysis, we stress substantively-theoretically guided (i.e., structural) specifications that can support counterfactual analyses of spatial or spatiotemporal responses in dependent-variable terms and that can distinguish possible sources of spatial association: common exposure, contagion, and selection.

Existing *S-QualDep* literature, excepting *S-Probit* work, leans Bayesian, CAR (conditional autoregressive) rather than SAR (simultaneous autoregressive), and spatial-statistical in tradition. Our approach is simpler and more structural; we seek estimators for the structural parameters of theoretically derived *S-QualDep* regression models, which will allow estimates of the effects of counterfactual changes in explanatory variables, based on the estimated parameters of these models, in terms of the outcomes of interest, and that will also yield estimates of the uncertainty of these parameter and effect estimates. The proposal illustrates with simple such models for *S-Poisson* count and *S-Exponential* duration models, and shows an example *S-Probit* counterfactual in terms of estimated probabilities with associated certainty estimates (and not in parameter or latent-variable terms as in existing work). The proposed work will refine these and develop other *S-Count*, *S-Duration*, and *S-QualDep* structural models; compare their small-sample properties and sensitivity to misspecification/assumption-violation with extant less-structural approaches and the non-spatial and naïve-spatial models that dominate applied work.

Likewise, currently developing spatial-econometric and spatial-statistical approaches to estimated-**W** models tend inductive and non/semi-parametric and therefore are relatively unhelpful for addressing our core questions and desired counterfactuals. Extant network-analytic *coevolution* models correspond to estimated- and potentially endogenous-**W** models, and do address these questions centrally. However, they also condition on lagged observations of outcomes (**y**) and networks (**W**) and the first observations, thereby requiring large-T samples and not allowing estimation of **y** and **W** fully simultaneously. Our approach is again simpler and more structural than the spatial-statistical/econometric ones, and it does allow estimation of fully simultaneous **y** and **W**. The proposal illustrates with our initial exploration of the m-STAR model—i.e., the spatiotemporal-lag model with multiple $\mathbf{W}_r$—showing its equivalence to a model where the strength and pattern of interdependence, $\rho\mathbf{W}$, is modeled as (linear-additive) function of these $\mathbf{W}_r$. The case where some $\mathbf{W}_r$ is a function of the outcome, **y**, corresponds to an endogenous-**W** model, and we sketch an estimator that combines spatial-instrumentation to estimate these endogenous $\mathbf{W}_r$ with the conditional or unconditional m-STAR likelihood. The proposed work will introduce m-STAR as an approach to estimated-**W** models, develop qualitative and nonlinear estimated-**W** models in this framework, develop and evaluate the proposed S-GMM within S-ML estimator of fully simultaneous coevolution models, explore small-sample properties and sensitivity to misspecification & assumption-violation of this and extant network-analytic and less-structural spatial-statistical/econometric approaches, and develop tools for calculating and presenting counterfactual spatiotemporal dynamics and effects for these models.