Lasso estimation of spatial weight matrices

Akhil Rao

May 5, 2016

This essay looks at two penalized regression estimators that are currently being developed to estimate spatial weight matrices: the Pooled Lasso estimator ([7]) and the IVLASSO estimator ([6]). The estimators are interesting because they allow researchers to model spatial interactions without requiring the researcher to supply a structure for the spatial weight matrices from prior knowledge.

The first estimator, the Pooled Lasso, is suited to estimating spillover effects occurring because of agent-specific characteristics. This may be useful for scenarios where a treatment has externalities, but seems less suited to modeling game-theoretic applications where the outcome variables the researcher is interested in are dependent on each other.

The second estimator, the IVLASSO, is more general but also more limited in the sense of only being asymptotically sign-consistent. This estimator allows for spillover effects from agents' characteristics as well as their outcomes, and can even handle lagged outcomes as characteristics with spillover effects. However, this comes at a cost - elements of the spatial weight matrix which are "small" will be incorrectly estimated as 0, while elements of the weight matrix which are "large" will be estimated with the correct sign but may not converge to the true values. The authors refer to this property as "partial sign-consistency". To make this estimator more useful, they write a model with spatial fixed effects separated from the spatial weight matrix, and show that the spatial fixed effects are accurate with high probability.

1 Technology review

1.1 Lasso estimators

Lasso ("Least Angle Shrinkage and Selection Operator") estimators are a form of penalized regression. They are useful for estimating high-dimensional models, such as cases where the number of parameters k is larger than the number of observations n. The Lasso estimator was first introduced by Tibshirani in 1996 [11]. It involves a minimization problem of the form:

$$\min_{\beta,\omega} Q(\beta,\omega) + \lambda||\omega||_1$$

where $Q()$ is an objective function like $(y - X_1\beta - X_2\omega)^2$ to be minimized, β and ω are $k_1 \times 1$ and $k_2 \times 1$ vectors of regression coefficients, ω are the regression coefficients whose dimension the researcher wishes to control, $||\omega||_1$ is the L_1-norm of ω, and λ is a penalty parameter that the researcher must tune. The higher λ is, the greater the penalty imposed on the coefficients in ω.

For predictive applications in fields such as genetics, λ is typically tuned through an iterative cross-validation procedure where the λ that is chosen is the one that minimizes the prediction error over the cross-validation process. For the models considered in this essay, such a procedure to choose λ is arguably not so useful if it results in inconsistency in parameter estimates ([9], [13]).

The "Adaptive Lasso", introduced by Zou in 2006 [13], generalizes the Lasso estimator by allowing the individual coefficients to have different penalty weights which are estimated from the data. The minimization problem is of the form:

$$\min_{\beta,\omega} Q(\beta,\omega) + \lambda \sum_{i=1}^{k_2} w_i |\omega_i|$$

When $\omega_i = 0$, the weight $w_i \to \infty$, while when $\omega_i \neq 0$, $w_i \to C < \infty$. The advantage of the Adaptive Lasso over the standard Lasso is that the Lasso's variable selection is inconsistent [13] while the Adaptive Lasso asymptotically selects the correct model.

Because of the kink in the absolute value function, Lasso estimators will result in a number of coefficients in ω being estimated as 0s. This is useful when the number of possible regressors is large, but the researcher believes the underlying data-generating process to be sparse in these regressors. Formally, this could be represented by a condition like $\sum_{i=1}^{k_2} \mathbb{I}(\omega_i \neq 0) = s << n$.

Lasso estimators are typically computed using the Least Angle Regression (LARS) algorithm described by Efron et. al [3].

1.2 Physical dependence measure

In a 2005 paper [12], Wu defines a set of dependence measures that are "quite different from strong mixing conditions", but facilitate the analysis of nonlinear time series, kernel estimators, and dependent stochastic processes. The key definitions that we look at here is Wu's "functional" or "physical" dependence measure and the notion of "p-stability", in contrast to "probabilistic" dependence measures such as strong mixing conditions. Wu's dependence measure makes use of the notion of coupling, described below. In what follows, I use Wu's notation.

Let ϵ_i, $i \in \mathbb{Z}$ be IID random variables and g be a measurable function such that $X_i = g(\dots,\epsilon_{i-1},\epsilon_i)$ is a stationary process. Let (ϵ'_i) be an IID copy of (ϵ_i), that is, (ϵ'_i) is a sequence where each ϵ'_i has the same distribution as ϵ_i. Define the shift processes $\xi = (\dots,\epsilon_{i-1},\epsilon_i)$ and $\xi' = (\dots,\epsilon'_{i-1},\epsilon'_i)$. For a set $I \subset \mathbb{Z}$, let $\epsilon_{j,I} = \epsilon'_j$ if $j \notin I$, and let $\xi_{i,I} = (\dots,\epsilon_{i-1,I},\epsilon_{i,I})$, and let $\xi*_i = \xi_{i,\{0\}}$. Then $\xi_{i,I}$ is a coupled version of ξ_i with ϵ_j replaced by ϵ'_j if $j \in I$.

Definition (Physical dependence measure): For $p > 0$ and $I \subset \mathbb{Z}$ let $\delta_p(I,n) = ||g(\xi_n) - g(\xi_{n,I})||_p$ and $\delta_p(n) = ||g(\xi_n) - g(\xi^*_n)||_p$. I refer to $\delta_p(n,I)$ as "physical dependence coefficients".

Definition (p-stability): Let $p \geq 1$. The process (X_n) is said to be p-strong stable if $\Delta_p := \sum_{n=0}^{\infty} \delta_p(n) < \infty$

For comparison, the strong mixing coefficient in this setting would be $\alpha_n = \sup\{|P(A \cup B) - P(A)P(B)| : A \in \mathcal{A}^0_{-\infty}, B \in \mathcal{A}^\infty_n\}$, where the sigma algebra $\mathcal{A}^n_m = \sigma(X_m,\dots,X_n)$, for $m \leq n$. If

$\alpha_n \to 0$, then (X_i) is said to be strong mixing.

The coupling allows us to compare the way in which the realizations of the process (X_i) change when one member of the sequence of innovations is replaced by another drawn from the same distribution. The physical dependence measure is this comparison carried out in L_p-norm, i.e. the L_p distance between (X_n) and its coupling. The notion of p-strong stability is a finiteness restriction on the sum of the physical dependence coefficients $\delta_p(n, I)$. Requiring a process to be p-strong stable seems to require $\delta_p(n) \to 0$ as $n \to \infty$, similar to the requirement on α_n for strong mixing. In the paper considered here, Lam and Souza make use of the physical dependence coefficients with a condition that is similar to a p-stability requirement.

Wu's purpose in developing this physical dependence measure is (1) to make it easier for researchers to measure the dependence between time series, since mixing conditions require a supremum to be taken over sigma algebras, and (2) to facilitate analyses of commonly encountered processes that are not strong mixing, such as the Bernoulli shift process described in [1]. Wu also defines another such measure, the "predictive dependence" measure, which I will not discuss here.

2 Pooled Lasso

Of the two estimators this seems like the more straightforward one. Manresa [7] considers a model of the form:

$$y_{it} = \alpha_i + \beta_i x_{it} + \sum_{j \neq i} \gamma_{ij} x_{jt} + \mu_i z_{it} + \theta w_{it} + \delta_t + \epsilon_{it}$$

where y_{it} is an individual outcome, x_{it} is an individual characteristic which may generate spillovers, z_{it} are individual characteristics which do not generate spillovers, δ_t are time fixed-effects, and w_{it} is a vector of controls. Individuals are named $i = 1, \ldots, N$ and time periods are labeled $t = 1, \ldots, T$. The γ_{ij} are the "spillover effects" of interest, and comprise the elements of the spatial weight matrix. This model considers a single characteristic which generates spillovers.

This model allows the econometrician to estimate a spatial weight matrix on individual characteristics directly in an iterative two-step procedure. Although this does not deal with the class of spatial processes where a spatial weight matrix is applied to individual outcomes (SAR models), it is potentially useful for researchers interested in cases where individual outcomes are independent but individuals may still exert spillover effects on each other through their characteristics. Unlike SAR models, Pooled Lasso does not require any IVs.

The spillover effects in this model are time-invariant (γ_{ij}, not γ_{ijt}), and may be endogenous with respect to covariates and unobservables.

2.1 Estimation

Estimating the above model involves minimizing the following criterion:

$$(\hat{\alpha}, \hat{\beta}, \hat{\Gamma}, \hat{\theta}, \hat{\delta}) = \text{argmin}_{(\alpha,\beta,\Gamma,\theta,\delta)} \;\; Q(\alpha, \beta, \Gamma, \theta, \delta) + \sum_{i=1}^{N} \lambda_i \sum_{j=1, j \neq i} |\gamma_{ij}| \hat{\sigma}_j$$

where $Q = \sum_{i=1}^{N} Q_i$, $Q_i = \sum_{t=1}^{T}(y_{it} - \alpha_i - \beta_i x_{it} - \sum_{j=1,j\neq i}^{N} \gamma_{ij} x_{jt} - \theta w_{it} - \delta_t)^2$, and $\hat{\sigma_j}^2 = T^{-1}\sum_{t=1}^{T}(x_{jt} - \bar{x_j})^2$.

The idea is to pick some initial values of θ, β, and δ and then iterate over two steps until convergence: the first involves estimating N Lasso regressions for each individual i's time-series to identify the spillover effects of other individuals' characteristics on i, and the second involves estimating panel OLS regressions to estimate the effect of individual i's own characteristics on their outcome.

At step m, estimate the N Lasso regressions for each i's time series:

$$(\alpha_i^{(m)}, \gamma_i^{(m)}) = \text{argmin}_{(\alpha_i,\gamma_i)} \left\{ Q_i(\alpha_i, \beta_i^{(m)}, \gamma_i, \theta^{(m)}, \delta^{(m)}) + \lambda_i \sum_{j\neq i} |\gamma_{ij}| \right\}$$

At step $m+1$, use OLS to update the values of θ, β, and δ:

$$(\beta^{(m+1)}, \theta^{(m+1)}, \delta^{(m+1)}) = \text{argmin}_{(\beta,\theta,\delta)} \;\; Q(\alpha^{(m)}, \beta, \gamma^{(m)}, \theta, \delta)$$

This is a convex optimization problem, and should converge to the global minimum. Intuitively, the first stage uses the time series comovements between own outcomes and others' covariates to estimate spillover effects from others, which are then held constant in the second stage to estimate the effects of own-characteristics from the panel data.

2.2 Key assumptions and consistency

Single source of spillovers: The model as presented above involves a single characteristic x_{jt} which causes spillovers. Appendix A.1 briefly describes an extension of this model with lagged outcomes $y_{j,t-1}$ as a case of a second characteristic that may cause spillovers. This is not considered in detail in the paper, but it seems doable.

Sparse Eigenvalues: Let $X_i = (x_{1t}, \dots, x_{i-1,t}, x_{i+1,t}, \dots, x_{Nt})$, and $C_i = X_i' X_i$. For $1 \leq u \leq N-1$, the u-sparse minimal and maximal eigenvalues are defined as the minimal or maximal eigenvalues of any $u \times u$ submatrix of C_i. Formally,

$$\Phi_{min}(u)[C_i] = \min_{\delta\in\mathbb{R}^{(N-1)}, 1\leq||\delta||_0<u} \frac{\delta' C_i \delta}{||\delta||_2}$$

$$\Phi_{max}(u)[C_i] = \max_{\delta\in\mathbb{R}^{(N-1)}, 1\leq||\delta||_0<u} \frac{\delta' C_i \delta}{||\delta||_2}$$

where $||\delta||_0 = \sum_{k=1}^{N-1} \mathbb{I}\{\delta_i \neq 0\}$.

As N,$T \to \infty$, $0 < \kappa'' \leq \Phi_{min}(s_i \log T)[C_i] \leq \Phi_{max}(s_i \log T)[C_i] \leq \kappa' < \infty$, where κ'' and κ' don't depend on N or T, and $\sum_{j\neq i} \mathbb{I}\{\gamma_{ij} \neq 0\} = s_i = o(T)$.

The bounds in this assumption seem standard to ensure invertibility and get consistency of the estimator. The dependence on s_i makes the "sparseness" requirement on the weight matrix precise. This condition is defined in a 2009 paper by Meinhausen and Yu [8].

When $\lambda_i \propto \sigma_i \sqrt{T \log N}$, this condition bounds the error of $\hat{\gamma}_{ij}$ as $\sum_{j\neq i}(\hat{\gamma}_{ij} - \gamma_{ij}^*)^2 = O_p\left(s_i \frac{\log N}{T}\right)$, where γ_{ij}^* is the true value of the parameter.

Independent noise: The second main assumption is that the noise in estimation of the γ_{ij} is independent across individuals, i.e. $E((\hat{\gamma}_{ij} - \gamma^*_{ij})(\hat{\gamma}_{kl} - \gamma^*_{kl})) = 0$. This assumption is combined with the sparse eigenvalues assumption to obtain the rate of convergence of the estimator.

Consistent model selection: The elements of the weight matrix in this model are spillover effects of j's characteristics on i. Consistent estimation of this matrix involves correctly selecting the characteristics of j which affect i and correctly estimating the size of the effect. In the Pooled Lasso model, the characteristic of agent j that is selected agent i's outcome is the one with the highest R^2 in predicting agent i's outcomes. More precisely, for the case of a single source of spillover, the estimator of the source of spillover to i is

$$\hat{j}(i) = \text{argmax}_{k \neq i} R^2_k$$

where R^2_j is the R^2 of the time series regression of the outcome of agent i on the characteristics of agent j. The probability that the identity of the source of spillovers is not the true source of spillovers depends on the behavior of $\sup_{k \neq j(i)} |T^{-1} \sum_{t=1}^{T} x_{kt}\epsilon_{it}|$, and requires $\frac{\log N}{T} \to 0$.

Selecting the tuning parameter λ by minimizing cross-validated predictive error is not ideal for inferential purposes, so Manresa chooses λ to uniformly bound the noise in estimation $n^{-1} \sum_{i=1}^{n} X_i \epsilon_i$, where X_i is an observation and ϵ_i is an innovation. This approach was described in [2].

3 IVLASSO

Lam and Souza [6] consider a model of the form:

$$y_t = W^*_1 y_t + W^*_2 X_t \beta^* + \mu^* + \epsilon_t$$

where $y_t = (y_{t1}, \ldots, y_{tN})'$ is an $N \times 1$ vector of dependent time series variables, W^*_j for $j = 1, 2$ are the spatial weight matrices to be estimated, and X_t is an $N \times K$ matrix of covariates. μ^* is a constant vector of "spatial fixed effects".

Lam and Souza focus on cases where X_t and ϵ_t are correlated, but there exist instrumental variables B_t of size $N \times K$ for each t which are correlated with y_t and X_t but not ϵ_t. Consistent estimation for this model requires $E(X_t) = 0$, but this can be relaxed without loss of generality. Lam and Souza discuss this relaxation in the paper; I ignore it here for simplicity.

Stacking the T equations above into regression form we get

$$y = M_{\beta^*}\xi^* + \mu^* \otimes \mathbb{1}_T + \epsilon$$

where $y = vec\{(y_1, \ldots, y_T)'\}$, $M_{\beta^*} = (I_N \otimes (y_1, \ldots, y_T)', I_N \otimes \{(I_T \otimes \beta^{*'})(X_1, \ldots, X_T)'\})$, $\xi^* = (vec(W^*_1), vec(W^*_2))$, and $\epsilon = vec\{(\epsilon_1, \ldots, \epsilon_T)'\}$.

Let B be the optimal instrument defined from the B_t; the definition is omitted here. The augmented model is

$$B'y = B'M_{\beta^*}\xi^* + B'\epsilon$$

3.1 Estimation

This model needs 3 estimators, one each for ξ^*, β^*, and μ^*.

First, we define the Lasso estimator $\tilde{\xi}$:

$$\begin{aligned}
&\tilde{\xi} = \text{argmin}_\xi \ \frac{1}{2T}||B'y - B'M_{\beta(\xi)}\xi||^2 + \gamma_T||\xi||_1 \\
&\text{s.t.} \ \sum_{j\neq i}|w_{1,ij}| < 1, \sum_{j\neq i}|w_{2,ij}| < 1, \\
&\beta(\xi) = \{X'W_2^{\otimes\prime}W_2^{\otimes}B^\nu B^{\nu\prime}W_2^{\otimes\prime}W_2^{\otimes}X\}^{-1}X'W_2^{\otimes\prime}W_2^{\otimes}B^\nu B^{\nu\prime}W_2^{\otimes\prime}(I_{TN} - W_1^{\otimes})y^\nu
\end{aligned}$$

where $\bar{B} = T^{-1}\sum_{i=1}^T B_t$, $B^\nu = ((B_1 - \bar{B})', \ldots, (B_T - \bar{B})')'$, $X = (X_1', \ldots, X_T')'$, $y = (y_1', \ldots, y_T')'$, and for a matrix M, $M^{\otimes} = I_T \otimes M$.

This gives us a consistent estimator for β^* as $\tilde{\beta} = \beta(\tilde{\xi})$. However, this estimator for ξ^* is biased and not asymptotically sign-consistent without restrictive requirements. To overcome this issue, Lam and Souza use an Adaptive Lasso estimator,

$$\begin{aligned}
&\hat{\xi} = \text{argmin}_\xi \ \frac{1}{2T}||B'y - B'M_{\beta(\tilde{\xi})}\xi||^2 + \gamma_T'||\xi||_1 \\
&\text{s.t.} \ \sum_{j\neq i}|w_{1,ij}| < 1, \sum_{j\neq i}|w_{2,ij}| < 1
\end{aligned}$$

This Adaptive Lasso estimator $\hat{\xi}$ is similar to the Lasso estimator $\tilde{\xi}$, except that $\hat{\xi}$ uses $\tilde{\xi}$ to define $\beta(\tilde{\xi})$, and $\hat{\xi}$ uses a data-dependent tuning parameter γ_T'. γ_T' may be the same as γ_T when the true weight matrices are sparse, but may be different if the true weight matrices are only approximately sparse as defined in assumption M2. The solution to the above minimization problem is found by an modified block coordinate descent algorithm which I don't discuss here.

Finally, we get the estimator for μ^* as $\hat{\mu} = (I_N - \hat{W}_1)\bar{y}$ by reconstructing $\hat{W}_1$ from $\hat{\xi}$.

3.2 Key assumptions and consistency

Reduced form (Assumption M1): There exists a constant $\eta > 0$ such that $||W_1^*||_\infty < \eta < 1$ and $||W_2^*||_\infty < \eta < 1$ uniformly as $N \to \infty$. This assumption assures invertibility of $(I_N - W_1^*)$ and guarantees the existence of the reduced form

$$\begin{aligned}
y_t &= \Pi_1^*\mu^* + \Pi_1^*W_2^*X_t\beta^* + \Pi_1^*\epsilon_t, \\
\Pi_1^* &= (I_N - W_1^*)^{-1}
\end{aligned}$$

Elements of the spatial weight matrices: Define the index sets

$$\begin{aligned}
J_0 &= \{j : \xi_j^* = 0 \text{ for all off-diagonal elements of } W_1^* \text{ and } W_2^*\} \\
J_1 &= \{j : |\xi_j^*| \geq \tau \text{ for all off-diagonal elements of } W_1^* \text{ and } W_2^*\} \\
J_2 &= \{j : 0 < |\xi_j^*| < \tau \text{ for all off-diagonal elements of } W_1^* \text{ and } W_2^*\}
\end{aligned}$$

Approximate sparseness/partial sign consistency (Assumption M2): There exists a constant $\tau > 0$ such that the elements $w^*_{1,ij}$ and $w^*_{2,ij}$ of the spatial weight matrices are constants as $N \to \infty$ whenever they are larger than or equal to τ in magnitude. For elements smaller than τ, we have $w^*_{1,ij}, w^*_{2,ij} \to 0$ as $N \to \infty$. The idea of "approximate sparseness" here is that if there is a constant τ such that the absolute sum of elements indexed by J_2 is somehow "small", then a partially sign-consistent estimator should get the sign right for the elements of J_1, while estimating the elements of J_0 and J_2 as 0. The partial sign consistency is proved in Theorem 4.

Physical dependence coefficients: Denote $\{b_t\} = \{vec(B_t)\}$, $\{x_t\} = \{vec(X_t)\}$, each are $NK \times 1$ vectors. We can see these as being defined by

$$x_t = [f_j(\mathcal{F}_t)]_{1\le j\le NK}, \quad b_t = [g_j(\mathcal{G}_t)]_{1\le j\le NK}, \quad e_t = [h_j(\mathcal{H}_t)]_{1\le j\le NK}$$

where $f_j()$, $g_j()$, and $h_j()$ are measurable functions defined on the real line, and the shift processes $\mathcal{F}_t = (\dots, e_{x,t-1}, e_{x,t})$, $\mathcal{G}_t = (\dots, e_{b,t-1}, e_{b,t})$, and $\mathcal{H}_t = (\dots, e_{\epsilon,t-1}, e_{\epsilon,t})$ are defined by the IID processes $\{e_{x,t}\}$, $\{e_{b,t}\}$, and $\{e_{\epsilon,t}\}$ respectively. Here, $\{e_{b,t}\}$ is correlated with $\{e_{x,t}\}$ but independent of $\{e_{\epsilon,t}\}$, so that $\{B_t\}$ are proper instruments for $\{X_t\}$. Now, for $d > 0$, define

$$\begin{aligned} \theta^x_{t,d,j} &= ||x_{tj} - x'_{tj}||_d \\ \theta^b_{t,d,j} &= ||b_{tj} - b'_{tj}||_d \\ \theta^\epsilon_{t,d,j} &= ||\epsilon_{tj} - \epsilon'_{tj}||_d \end{aligned}$$

where $x'_{tj} = f_j(\mathcal{F}'_t)$, $\mathcal{F}'_t = (\dots, e_{x,t-1}, e'_{x,0}, e_{x,t+1}, \dots)$ is a coupled version of x_{tj} with $e_{x,0}$ replaced with an IID copy $e'_{x,0}$.

Time dependence (Assumption T1): Define

$$\begin{aligned} \Theta^x_{m,a} &= \sum_{t=m}^{\infty} \max_{1\le j\le NK} \theta^x_{t,d,j}, \\ \Theta^b_{m,a} &= \sum_{t=m}^{\infty} \max_{1\le j\le NK} \theta^b_{t,d,j}, \\ \Theta^\epsilon_{m,a} &= \sum_{t=m}^{\infty} \max_{1\le j\le NK} \theta^\epsilon_{t,d,j} \end{aligned}$$

Then for $w > 2$, $\Theta^x_{m,2w}, \Theta^b_{m,2w}, \Theta^\epsilon_{m,2w} \le Cm^{-\alpha}$, with $\alpha, C > 0$ being constants that can depend on w.

These conditions are similar to the p-stability requirement Wu associates with the physical dependence coefficients. Stationary Markov chains and stationary linear processes are examples of time series that satisfy T1.

This condition, combined with a tail condition in assumption D3 that provides an exponential bound on $P(|Z| > v)$ for $B_{t,jk}$, $X_{t,jk}$, and $\epsilon^*_{t,j}$, is used in Lemma 1 to construct an inequality similar to the Fuk-Nagaev inequality ([10], [5]) that bounds $P(|T^{-1}\sum_{t=1}^{T} x_{tj}| > v)$ and is useful in proving consistency of $\tilde{\beta}$ and the asymptotic sign-consistency of $\hat{\xi}$. It is a Nagaev-type inequality in the sense that the bound is the sum of a linearly shrinking term and an exponentially shrinking term.

Other contributions of this paper which I have not discussed include two BIC criteria to select the appropriate penalization parameters for the Lasso and Adaptive Lasso problems and the modified block coordinate descent algorithm to solving the Adaptive Lasso problem.

4 Conclusion

The Pooled Lasso estimator seems much more straightforward than the IVLASSO estimator, albeit less flexible in modeling spillover effects. The cost to this flexibility is that the IVLASSO is not consistent in the traditional sense; small sources of spillovers will be incorrectly estimated as 0, while large sources of spillovers will be estimated with the correct sign but not necessarily the correct value. The use of consistently estimated spatial fixed effects in IVLASSO remedies this to some extent.

It is difficult to compare the two estimators directly, but it seems that Pooled Lasso is easier to apply. The conditions on the physical dependence coefficients of the IVLASSO estimator seem difficult to check in practice. Lam and Souza present the results of applying IVLASSO to estimating the heterogeneity of the price response in different sectors of the economy to unexpected macroeconomic shocks. However, I was unable to find details on how they implemented this, specifically the choice of instrumental variables B_t and how the conditions of assumption T1 were checked.

It seems that estimating the spatial weight matrix directly with Lasso estimators can result in biased coefficient estimates ([13], [4]). Lam and Souza get around this by using an Adaptive Lasso estimator ([13]), which introduces some additional complexity in estimating the coefficient-specific weights. Manresa gets around this by using a non-cross-validation-based approach to choosing the penalization parameter to uniformly bound the estimation noise, as described in [2].

Moving forward, I think I can apply at least one of these models to estimate the spillover effects of water trading on agricultural output in Australia, using data on all water trades in the country between 2009-2013. Unobserved factors that affect the agricultural output in an area, such as soil quality or water trading by farmers whose land straddles two or more trading zones, may result in "spillover effects" of trading in one zone on the output in another zone.

A version of the model I am interested in is:

$$NDVI_{it} = \alpha_i + \beta_i trade_{it} + \sum_{j \neq i} \gamma_{ij} trade_{jt} + \mu_i z_{it} + \theta w_{it} + \delta_t + \epsilon_{it}$$

where $NDVI_{it}$ are satellite measurements of "vegetation greenness" (a proxy for agricultural output) for trading region i in month t, $trade_{it}$ is the sum of within-region flows and flows to all other regions for region i in month t, γ_{ij} are the spillover effects of interest, z_{it} are region-specific characteristics that do not cause spillovers such as the historical number of water entitlements and the amount of rainfall in the trading region, and δ_t are month fixed-effects.

One complication is that while the trading data is daily, the satellite images offer only monthly data and there are around 90 water trading regions. Practically, the time dimension is less than 50 (monthly observations for 4 years), while the number of potential cross-sectional spillovers is on the order of 90×90. Pooled Lasso seems to be a useful tool for this setting.

My dataset also includes transaction prices, and IVLASSO may be useful to estimate water demand and supply in each trading region incorporating spillover effects of other regions' prices and trades.

References

[1] Donald W K Andrews. Non-strong mixing autoregressive processes. *Journal of Applied Probability*, 21:930–934, 1984.

[2] PJ Bickel, Y Ritov, and A Tsybakov. Simultaneous analysis of lasso and dantzig selector. *Annals of Statistics*, 37:1705–1732, 1979.

[3] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of Statistics*, 32:407–499, 2004.

[4] Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 2001.

[5] D H Fuk and S V Nagaev. Probability inequalities for sums of independent random variables. *Theory of Probability and its Applications*, 16:643–660, 2006.

[6] Clifford Lam and Pedro CL Souza. One-step regularized spatial weight matrix and fixed effects estimation with instrumental variables. 2016.

[7] Elena Manresa. Estimating the structure of social interactions using panel data. 2015.

[8] N Meinhausen and B Yu. Lasso-type recovery of sparse representations for high-dimensional data. *Annals of Statistics*, 37:2246–2270, 2006.

[9] Nicolai Meinshausen and Peter Buhlmann. Variable selection and high-dimensional graphs with the lasso. *The Annals of Statistics*, 34:1436–1462, 2006.

[10] S V Nagaev. Large deviations of sums of independent random variables. *Annals of Probability*, 7:745–789, 1979.

[11] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of Royal Statistical Society Series B*, 58:267–288, 1996.

[12] Wei Biao Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 102:14150–14154, 2005.

[13] Hui Zou and Hao Helen Zhang. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101:1418–1429, 2006.