ONLINE APPENDIX

BENJAMIN J. GILLEN, SERGIO MONTERO, HYUNGSIK ROGER MOON AND MATTHEW SHUM

APPENDIX A: ITERATIVE COMPUTATION OF PENALTY LOADINGS

Our penalized estimators apply data-dependent factor loadings for each of the coefficients included in the model. The data-dependent factor loadings scale the penalty for each coefficient according to the variability of the associated coefficient and the model residual. These loadings appear in Υ_β and Υ_ω in Algorithm 5.1, Υ_θ in Equation (5.5), Υ_ϕ in Equation (5.11), and Υ_ζ in Equation (5.12). Here, we review the application of the iterative approach of Belloni et al. (2013) to computing these penalty loadings.

A.1. ITERATIVE COMPUTATION FOR LINEAR MODELS

Recalling the formula for Step I of Algorithm 5.1:

$$\min_{\beta\in\mathbb{R}^{K_T+1}} \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\mathcal{S}_{jt}-x_{0t}'\beta_{0j}-x_{1jt}'\beta_1-p_{jt}\beta_p\right)^2+\frac{\lambda_\beta}{T}\|\hat{\Upsilon}_\beta\beta\|_1.$$

We set $\lambda_\beta = 2c\sqrt{JT}\Phi^{-1}\left(1-\gamma/2(K_T+1)\right)$, with the Belloni et al. (2013) recommended values being $c = 1.1$ and $\gamma = 0.05/\log\left(K_T+1\vee T\right)$. The k^{th} diagonal entry in $\hat{\Upsilon}_\beta$ scales the penalty according to the variability in the k^{th} regressor, which we will denote $x_{k,jt}$, and the residual $\epsilon_{jt} \equiv \mathcal{S}_{jt}-x_{0t}'\beta_{0j}-x_{1jt}'\beta_1-p_{jt}\beta_p$. The infeasible ideal sets $\hat{\Upsilon}_{\beta,\{k,k\}} = \sqrt{\mathbb{E}\left[x_{k,jt}^2\epsilon_{jt}^2\right]}$. The iterative Algorithm A.1 initializes Υ_β with the expected squared value of each regressor, fits the LASSO regression, recovers the residuals, and uses these residuals to compute the sample analog to the ideal value. This algorithm extends immediately to Υ_ω. Defining the residual $\varepsilon_{jt} \equiv p_{jt}-x_{0t}'\omega_{0j}-x_{1jt}'\omega_1$, the infeasible ideal penalty values for this problem are $\hat{\Upsilon}_{\omega,\{k,k\}} = \sqrt{\mathbb{E}\left[x_{k,jt}^2\varepsilon_{jt}^2\right]}$. For completeness, the calculation is detailed in Algorithm A.2.

Algorithm A.1 Iterative Algorithm for Υ_β

I. Initialize $\Upsilon_{k,k}^0 = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x_{k,j,t}^2}$, $k = 1,\ldots,K_T$.

II. For $\mathcal{I} = 1, \ldots, \bar{\mathcal{I}}$, or until $\|\Upsilon^{\mathcal{I}}-\Upsilon^{\mathcal{I}-1}\| < \delta$:

a) Solve $\hat{\beta} = \arg\min_{\beta\in\mathbb{R}^{K_T+1}} \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\mathcal{S}_{jt}-x_{0t}'\beta_{0j}-x_{1jt}'\beta_1-p_{jt}\beta_p\right)^2+\frac{\lambda_\beta}{T}\|\Upsilon^{\mathcal{I}-1}\beta\|_1.$

b) Compute the residuals: $\hat{\epsilon}_{jt} \equiv \mathcal{S}_{jt}-x_{0t}'\hat{\beta}_{0j}-x_{1jt}'\hat{\beta}_1-p_{jt}\hat{\beta}_p$.

c) Update $\Upsilon_{k,k}^{\mathcal{I}} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x_{k,j,t}^2\hat{\epsilon}_{jt}^2}$, $k = 1,\ldots,K_T$.

III. Set $\hat{\Upsilon}_\beta = \Upsilon^{\mathcal{I}}$.

Algorithm A.2 Iterative Algorithm for Υ_ω

I. Initialize $\Upsilon^0_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}}$, $k = 1, \ldots, K_T$.

II. For $\mathcal{I} = 1, \ldots, \bar{\mathcal{I}}$, or until $\|\Upsilon^{\mathcal{I}} - \Upsilon^{\mathcal{I}-1}\| < \delta$:

a) Solve $\hat{\omega} = \min_{\omega\in\mathbb{R}^{K_T}} \dfrac{1}{JT}\displaystyle\sum_{t=1}^{T}\sum_{j=1}^{J}\left(p_{jt} - x'_{0t}\omega_{0j} - x'_{1jt}\omega_1\right)^2 + \frac{\lambda_\omega}{T}\|\hat{\Upsilon}_\omega\omega\|_1$.

b) Compute the residuals: $\hat{\varepsilon}_{jt} \equiv p_{jt} - x'_{0t}\hat{\omega}_{0j} - x'_{1jt}\hat{\omega}_1$.

c) Update $\Upsilon^{\mathcal{I}}_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}\hat{\varepsilon}^2_{jt}}$, $k = 1, \ldots, K_T$.

III. Set $\hat{\Upsilon}_\omega = \Upsilon^{\mathcal{I}}$.

A.2. ITERATIVE COMPUTATION FOR NONLINEAR MODELS

The selection in nonlinear models requires accounting for the additional estimation error introduced by selection on a generated regressor. Consequently, the residual with which to scale the regressor's variability must be augmented by the variance of the generated selection target. Recall the selection problem in Equation (5.11):

$$\tilde{\phi} = \arg\min_{\phi\in\mathbb{R}^{K_T}} \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\tilde{\delta}_{jt} - x'_{0t}\phi_{0j} - x'_{1jt}\phi_1\right)^2 + \frac{\lambda_\phi}{T}\|\hat{\Upsilon}_\phi\phi\|_1.$$

The Υ_ϕ matrix requires a slight adjustment to account for estimation error in the $\tilde{\delta}_{jt}$s. Defining

$$\epsilon_{\delta,jt} \equiv \delta_{jt} - \tilde{\delta}_{jt} = \tilde{\delta}_{jt} = \tilde{x}'_{jt}\left(\tilde{\beta}_j - \beta_j\right) + x'_{1jt}\left(\tilde{\beta}_1 - \beta_1\right) + p_{jt}\left(\tilde{\beta}_p - \beta_p\right) + \tilde{\xi}_{jt} - \xi_{jt}$$

and $\epsilon_{\phi,jt} \equiv \tilde{\delta}_{jt} - x'_{0t}\phi_{0j} - x'_{1jt}\phi_1$, the ideal weight for $\phi_{0j,k}$ is equal to $\sqrt{\bar{E}\left[x^2_{0t,k}\left(\epsilon_{\delta,jt} + \epsilon_{\phi,jt}\right)^2\right]}$ and to $\sqrt{\bar{E}\left[x^2_{1jt,k}\left(\epsilon_{\delta,jt} + \epsilon_{\phi,jt}\right)^2\right]}$ for $\phi_{1,k}$.

We can define $\bar{x}_{jt} = \left[\tilde{x}'_{jt}x'_{1jt}, p_{jt}\right]'$ and Σ_j as the rows and columns of the variance-covariance matrix for $\tilde{\beta}$ computed using the sandwich covariance matrix from the solution to (5.10):

$$\hat{\epsilon}^2_{\delta,jt} = E\left[\epsilon^2_{\delta,jt}\right] = \bar{x}'_{jt}\Sigma_j\bar{x}_{jt} + \sigma^2_\xi.$$

For feasible implementation, we again initialize the Υ_ϕ matrix with the diagonal variances of the regressors. We then recursively solve (5.11) to recover the residuals $\epsilon_{\phi,jt}$ and update the Υ_ϕ accordingly.

The approach above does not apply as readily to the solution for (5.12), as we cannot easily characterize the variance of the optimum instruments for the nonlinear features of the model. However, we do not need to account for the population variance of the asymptotic optimal instruments in our selection of controls. Importantly, the estimated optimal instruments provide the only source of exogenous variation used to identify the heterogeneity in voter impressionability. Consequently, performing selection on the utilized instruments as if they represented the population optimal instruments suffices to control for observable heterogeneity. Recalling the penalization problem:

$$\tilde{\zeta} = \arg\min_{\zeta\in\mathbb{R}^{K_T}} \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\tilde{z}_{v,jt} - x'_{0t}\zeta_{0j} - x'_{1jt}\zeta_1\right)^2 + \frac{\lambda_\zeta}{T}\|\hat{\Upsilon}_\zeta\zeta\|_1$$

and defining the residual $\varepsilon_\zeta = \tilde{z}_{v,jt} - x'_{0t}\zeta_{0j} - x'_{1jt}\zeta_1$, the ideal $(k, k)^{th}$ entry in $\Upsilon_\zeta = \mathbb{E}\left[x^2_{k,jt}\varepsilon^2_\zeta\right]$. We can then apply the approach from Algorithms A.1 and A.2.

Algorithm A.3 Iterative Algorithm for Υ_ϕ

I. Initialize $\Upsilon^0_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}}$, $k = 1, \ldots, K_T$.

II. Compute $\hat{\epsilon}^2_{\delta,jt} = \bar{x}'_{jt}\Sigma_j\bar{x}_{jt} + \sigma^2_\xi$ from the solution to the feasible GMM problem (5.10).

III. For $\mathcal{I} = 1, \ldots, \bar{\mathcal{I}}$, or until $\|\Upsilon^{\mathcal{I}} - \Upsilon^{\mathcal{I}-1}\| < \delta$:

a) Solve $\tilde{\phi} = \arg\min_{\phi\in\mathbb{R}^{K_T}} \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\tilde{\delta}_{jt} - x'_{0t}\phi_{0j} - x'_{1jt}\phi_1\right)^2 + \frac{\lambda_\phi}{T}\|\Upsilon^{\mathcal{I}-1}\phi\|_1$.

b) Compute the residuals: $\hat{\epsilon}_{\phi,jt} \equiv \tilde{\delta}_{jt} - x'_{0t}\hat{\phi}_{0j} - x'_{1jt}\hat{\phi}_1$.

c) Update $\Upsilon^{\mathcal{I}}_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}\left(\hat{\epsilon}^2_{\phi,jt} + \hat{\epsilon}^2_{\delta,jt}\right)}$, $k = 1, \ldots, K_T$.

IV. Set $\hat{\Upsilon}_\phi = \Upsilon^{\mathcal{I}}$.

Algorithm A.4 Iterative Algorithm for Υ_ζ

I. Initialize $\Upsilon^0_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}}$, $k = 1, \ldots, K_T$.

II. For $\mathcal{I} = 1, \ldots, \bar{\mathcal{I}}$, or until $\|\Upsilon^{\mathcal{I}} - \Upsilon^{\mathcal{I}-1}\| < \delta$:

a) Solve $\hat{\zeta} = \arg\min_{\zeta\in\mathbb{R}^{K_T}} \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\tilde{z}_{v,jt} - x'_{0t}\zeta_{0j} - x'_{1jt}\zeta_1\right)^2 + \frac{\lambda_\zeta}{T}\|\hat{\Upsilon}_\zeta\zeta\|_1$.

b) Compute the residuals: $\hat{\varepsilon}_{\zeta,jt} = \tilde{z}_{v,jt} - x'_{0t}\hat{\zeta}_{0j} - x'_{1jt}\hat{\zeta}_1$.

c) Update $\Upsilon^{\mathcal{I}}_{k,k} = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{JT} x^2_{k,j,t}\hat{\varepsilon}^2_{\zeta,jt}}$, $k = 1, \ldots, K_T$.

III. Set $\hat{\Upsilon}_\zeta = \Upsilon^{\mathcal{I}}$.

A.3. GMM PENALTY FOR VERIFYING FIRST-ORDER CONDITIONS

While we do not directly evaluate the objective function in the global parameter space for Equation (5.5), we do need to verify the first-order conditions for the local solution based on the selected model in the last step of Algorithm 5.2:

$$q_k \equiv \frac{\partial}{\partial\beta_{0jk}} Q\left(\tilde{\theta}^*, \tilde{x}^k, z, p, s\right) < \lambda_\theta \upsilon_k, k = 1, \ldots, K_0, j = 1, \ldots, J.$$

As discussed in the text, the infeasible ideal value of $\upsilon_k = \sqrt{\bar{E}\left[x^2_{0t,k}\xi^2_{jt}\right]}$. Here, we are already working from a (putative) local optimum, so we can take the estimated values $\xi\left(\tilde{\theta}^*, \tilde{x}, z, p, s\right)$ to estimate the empirical analog to the expectation:

$$\hat{\upsilon}_k = \sqrt{\frac{1}{JT}\sum_{j,t=1}^{J,T} x^2_{0t,k}\tilde{\xi}^2_{jt}}.$$

This calculation has the added benefit of being computable variable-by-variable to mitigate memory and computational limitations.

REFERENCES-APPENDIX

Belloni, A., V. Chernozhukov, and C. Hansen (2013). Inference on treatment effects after selection amongst high-dimensional controls. *Review of Economic Studies 81, 608-50.* 81, 608-50.