Appendix

A Additional Illustration for Section 2

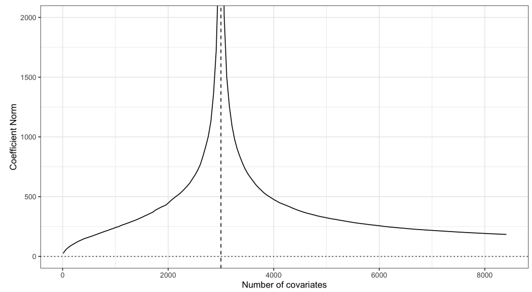

Figure 7: Average of the coefficient norm $\|\hat{\beta}^J\|$ for varying size of the set J of covariates.

As a further illustration of the behavior of linear regression in Figure 1 as complexity increases, Figure 7 shows the average norm of the model coefficients across different model complexities. Starting small, model coefficients initially grow (in terms of their Euclidean norm), reaching their peak at the interpolation threshold. To the right of the interpolation threshold, the norm of the model coefficients decreases mechanically, since the estimator now minimizes the norm among all interpolating solutions with fewer and fewer sparsity constraints.

B Proofs

Proof of Proposition 1. We provide a direct proof for the choice $\lambda_j = \frac{1-X_j'(X_J X_J')^{-1}X_j}{|J|-n}$ via the Sherman–Morrison–Woodbury formula. We first note that, for $k \geq n$, $A \in \mathbb{R}^{n\times k}$ of full rank n, $a \in \mathbb{R}^n$, and $\hat{\alpha} = \arg\min_{\alpha\in\mathbb{R}^k; A\alpha=a} \|\alpha\|$ we have that $\hat{\alpha} = A'(AA')^{-1}a$. Indeed, $A\hat{\alpha} = a$, and for any $\alpha \in \mathbb{R}^k$ with $A\alpha = a$ and $\alpha \neq \hat{\alpha}$, for $\Pi = A'(AA')^{-1}A$ we have that

$$\|\alpha\|^2 = \|\Pi\alpha\|^2 + \|(\mathbb{I}-\Pi)\alpha\|^2 = \|\Pi\hat{\alpha}\|^2 + \|(\mathbb{I}-\Pi)(\alpha-\hat{\alpha})\|^2 = \|\hat{\alpha}\|^2 + \|\alpha-\hat{\alpha}\|^2 > \|\hat{\alpha}\|^2.$$

We next write $X^J \in \mathbb{R}^{n\times k}$ for the matrix with columns $X_j^J = X_j$ for $j \in J$ and $X_j^J = \mathbf{0}$ for $j \notin J$. Applying the above result to $\hat{\beta}^J$ and $\hat{\beta}^{J\setminus\{j\}}$ for all $j \in J$, we find

$$\hat{\beta}^J = X^{J\prime}(X_J X_J')^{-1}Y, \qquad \hat{\beta}^{J\setminus\{j\}} = X^{J\setminus\{j\}\prime}(X_{J\setminus\{j\}}X_{J\setminus\{j\}}')^{-1}Y.$$

Using that $X_{J\setminus\{j\}}X_{J\setminus\{j\}}' = X_J X_J' - X_j X_j'$, which is invertible by the assumption that $X_{J\setminus\{j\}}$ is of full rank, we find by the Sherman–Morrison–Woodbury that

$$(X_{J\setminus\{j\}}X_{J\setminus\{j\}})^{-1} = (X_J X_J')^{-1} + (X_J X_J')^{-1}X_j(1-X_j'(X_J X_J')^{-1}X_j)^{-1}X_j'(X_J X_J')^{-1} \quad (4)$$

with $X_j'(X_J X_J')^{-1}X_j \neq 1$. Plugging in,

$$\begin{aligned}
\hat{\beta}^{J\setminus\{j\}} &= (X^J - X^{\{j\}})'(X_{J\setminus\{j\}}X_{J\setminus\{j\}})^{-1}Y \\
&= \hat{\beta}^J - X^{\{j\}\prime}(X_J X_J')^{-1}Y - X^{\{j\}\prime}(X_J X_J')^{-1}Y\frac{X_j'(X_J X_j')^{-1}X_j}{1-X_j'(X_J X_j')^{-1}X_j} \\
&\quad + X^{J\prime}(X_J X_J')^{-1}X_j X_j'(X_J X_J')^{-1}Y\frac{1}{1-X_j'(X_J X_j')^{-1}X_j} \\
&= \hat{\beta}^J + \left(X^{J\prime}(X_J X_J')^{-1}X_j X_j'(X_J X_J')^{-1} - X^{\{j\}\prime}(X_J X_J')^{-1}\right)Y\frac{1}{1-X_j'(X_J X_j')^{-1}X_j}.
\end{aligned}$$

Since $\sum_{j\in J} X_j X_j' = X_J X_J'$ and $\sum_{j\in J} X^{\{j\}} = X^J$, we have that

$$\sum_{j\in J}\hat{\beta}^{J\setminus\{j\}}\lambda_j = \hat{\beta}^J\sum_{j\in J}\lambda_j + (|J|-n)(X^{J\prime}(X_J X_J')^{-1}X_J X_J'(X_J X_J')^{-1}Y - X^{J\prime}(X_J X_J')^{-1}Y) = \hat{\beta}^J\sum_{j\in J}\lambda_j.$$

Finally, $X_j'(X_J X_J')^{-1}X_j \geq 0$ since $X_J X_J'$ positive definite, $X_j'(X_J X_J')^{-1}X_j \leq 1$ since $X_{J\setminus\{j\}}X_{J\setminus\{j\}}' = X_J X_J' - X_j X_j' \preceq X_J X_J'$ in (4), and $\sum_{j\in J} X_j'(X_J X_J')^{-1}X_j = \operatorname{tr}\left(\sum_{j\in J} X_j X_j' X_J X_J'\right) = n$, so $\lambda_j \in [0,1]$ for all $j \in J$ and $\sum_{j=1}^J \lambda_j = 1$. $\square$

Proof of Proposition 2. The result follows from Proposition 3 by noting that, for two indpendent draws Y_A and Y_B for fixed X, we have that

$$\begin{aligned}\mathrm{E}[\|\hat\beta_A^J - \hat\beta_B^J\|^2|X] &= \mathrm{E}[\|(\hat\beta_A^J - \mathrm{E}[\hat\beta^J|X]) - (\hat\beta_B^J - \mathrm{E}[\hat\beta^J|X])\|^2|X] \\ &= \mathrm{E}[\|\hat\beta_A^J - \mathrm{E}[\hat\beta^J|X]\|^2|X] + \mathrm{E}[\|\hat\beta_B^J - \mathrm{E}[\hat\beta^J|X]\|^2|X] = 2\operatorname{tr}\operatorname{Var}(\hat\beta^J|X)\end{aligned}$$

(and the same for $J \setminus \{j\}$), and thus

$$\begin{aligned}\operatorname{tr}\operatorname{Var}(\hat\beta^J|X) &= \frac{1}{2}\mathrm{E}[\|\hat\beta_A^J - \hat\beta_B^J\|^2|X] \leq \frac{1}{2}\mathrm{E}\left[\min_{j\in J}\|\hat\beta_A^{J\setminus\{j\}} - \hat\beta_B^{J\setminus\{j\}}\|^2 \middle| X\right] \\ &\leq \min_j \frac{1}{2}\mathrm{E}[\|\hat\beta_A^{J\setminus\{j\}} - \hat\beta_B^{J\setminus\{j\}}\|^2|X] = \operatorname{tr}\operatorname{Var}(\hat\beta^{J\setminus\{j\}}|X). \qquad \square\end{aligned}$$

Proof of Proposition 3. Consider first the case $|J| \leq n$. Under Assumption 1, we define the projection matrices

$$\Pi^J = X_J(X_J'X_J)^{-1}X_J' \in \mathbb{R}^{n\times n}, \quad \Pi^{J\setminus\{j\}} = X_{J\setminus\{j\}}(X_{J\setminus\{j\}}'X_{J\setminus\{j\}})^{-1}X_{J\setminus\{j\}}' \in \mathbb{R}^{n\times n}.$$

Since $\Pi^{J\setminus\{j\}} = \Pi^{J\setminus\{j\}}\Pi^J$, we have that $X\hat\beta^{J\setminus\{j\}} = \Pi^{J\setminus\{j\}}Y = \Pi^{J\setminus\{j\}}\Pi^J Y = \Pi^{J\setminus\{j\}}\hat\beta^J$. Therefore,

$$\begin{aligned}\|\hat\beta_A^J - \hat\beta_B^J\|_{X'X}^2 = \|X\hat\beta_A^J - X\hat\beta_B^J\|^2 &= \|\Pi^{J\setminus\{j\}}(X\hat\beta_A^J - X\hat\beta_B^J)\|^2 + \|(\mathbb{I} - \Pi^{J\setminus\{j\}})(X\hat\beta_A^J - X\hat\beta_B^J)\|^2 \\ &\geq \|\Pi^{J\setminus\{j\}}(X\hat\beta_A^J - X\hat\beta_B^J)\|^2 = \|X\hat\beta_A^{J\setminus\{j\}} - X\hat\beta_B^{J\setminus\{j\}}\|^2 = \|\hat\beta_A^{J\setminus\{j\}} - \hat\beta_B^{J\setminus\{j\}}\|_{X'X}^2.\end{aligned}$$

Consider now the case $|J| > n$. Using the notation from the proof of Proposition 1, under Assumption 1 we have that $X^J\hat\beta^J = X\hat\beta^J = Y = X\hat\beta^{J\setminus\{j\}} = X^J\hat\beta^{J\setminus\{j\}}$ and thus $\Pi\hat\beta^J = \Pi\hat\beta^{J\setminus\{j\}}$ (as well as $(\mathbb{I} - \Pi)\hat\beta^J = \mathbf{0}$) for the projection matrix $\Pi = X^{J\prime}(X_J X_J')^{-1}X^J \in \mathbb{R}^{k\times k}$. As a consequence,

$$\begin{aligned}\|\hat\beta_A^J - \hat\beta_B^J\|^2 &= \|\Pi(\hat\beta_A^J - \hat\beta_B^J)\|^2 + \|(\mathbb{I} - \Pi)(\hat\beta_A^J - \hat\beta_B^J)\|^2 = \|\Pi(\hat\beta_A^{J\setminus\{j\}} - \hat\beta_B^{J\setminus\{j\}})\|^2 \\ &\leq \|\Pi(\hat\beta_A^{J\setminus\{j\}} - \hat\beta_B^{J\setminus\{j\}})\|^2 + \|(\mathbb{I} - \Pi)(\hat\beta_A^{J\setminus\{j\}} - \hat\beta_B^{J\setminus\{j\}})\|^2 = \|\hat\beta_A^{J\setminus\{j\}} - \hat\beta_B^{J\setminus\{j\}}\|^2. \; \square\end{aligned}$$

Proof of Proposition 4. Building upon the notation from Section 3.1, for $J \subset \{1,\ldots,N\}$ write $\mathcal{W}^J = \{w \in \mathcal{W}; w_j = 0 \text{ for all } j \notin J\}$ (where $\mathcal{W} = \{w \in [0,1]^N; \sum_{i=1}^N w_i = 1\}$ is the $N-1$-simplex) and let $\partial\mathcal{W}^J = \bigcup_{j\in J}\mathcal{W}^{J\setminus\{j\}} \subseteq \mathcal{W}^J$ be the boundary of $\mathcal{W}^J$. For outcomes, it will also be convenient to write $X = (y_{it})_{t\in\{1,\ldots,T\},i\in\{1,\ldots,N\}} \in \mathbb{R}^{T\times N}$ for the pre-treatment outcomes of the control units (with columns representing units), and $y = (y_{0t})_{t\in\{1,\ldots,T\}} \in \mathbb{R}^T$ for the pre-treatment outcomes of the treated unit.

As the first step, we note that we can express the quality of synthetic control weights $w \in \mathcal{W}^J$ as

$$\|Xw - y\|^2 = \|Xw - \overline{y}^J\|^2 + \|\overline{y}^J - y\|^2 \tag{5}$$

in terms of the fitted values $\overline{y}^J = X\overline{w}^J$ for the solution $\overline{w}^J$ to a relaxed problem that drops the non-negativity constraint. That solution with weights in $\mathcal{W}^* = \{w \in \mathbb{R}^N; \sum_{i=1}^n w_i = 1\}$ is defined, analogously to the synthetic-control solution in (1), as

$$\overline{w}^J = \operatorname*{arg\,min}_{w\in\overline{\mathcal{W}}^J}\|w\| \in \mathcal{W}^*, \qquad \overline{\mathcal{W}}^J = \operatorname*{arg\,min}_{w\in\mathcal{W}^*; w_j = 0 \forall j\notin J}\|Xw - y\| \subseteq \mathcal{W}^*.$$

For this solution, we note that $\|Xw - y\|^2 = \|X(w - \overline{w}^J)\|^2 + 2(w - \overline{w}^J)'X'(X\overline{w}^J - y) + \|X\overline{w}^J - y\|^2$. Assume now that $(w - \overline{w}^J)'X'(X\overline{w}^J - y) \neq 0$. Then there is some $\varepsilon \neq 0$ such that $\overline{w}^J(\varepsilon) = \overline{w}^J - (w - \overline{w}^J)\,\varepsilon \in \mathcal{W}^*$ with $w_j = 0$ for $j \notin J$ fulfills $\|X\overline{w}^J(\varepsilon) - y\| < \|X\overline{w}^J - y\|$, contradicting the choice of $\overline{w}^J$. Hence we must have that $\|Xw - y\|^2 = \|X(w - \overline{w}^J)\|^2 + \|X\overline{w}^J - y\|^2 = \|Xw - \overline{y}^J\|^2 + \|\overline{y}^J - y\|^2$, which establishes (5).

As the second step, we note that we can therefore define the synthetic control solution in (1) in terms of the fitted values $\overline{y}^J$ of the relaxed solution as

$$\hat{w}^J = \underset{w \in \widehat{\mathcal{W}}^J}{\arg\min} \|w\| \in \mathcal{W}^J, \qquad \widehat{\mathcal{W}}^J = \underset{w \in \mathcal{W}^J}{\arg\min} \|Xw - \overline{y}^J\| \subseteq \mathcal{W}^J.$$

This follows immediately from (5) since $\|\overline{y}^J - y\|$ is not affected by the choice of $w \in \mathcal{W}^J$. Similarly, for the constrained solutions with index set $J \setminus \{j\}$ for $j \in J$, we have that

$$\hat{w}^{J\setminus\{j\}} = \underset{w \in \widehat{\mathcal{W}}^{J\setminus\{j\}}}{\arg\min} \|w\| \in \mathcal{W}^{J\setminus\{j\}}, \qquad \widehat{\mathcal{W}}^{J\setminus\{j\}} = \underset{w \in \mathcal{W}^{J\setminus\{j\}}}{\arg\min} \|Xw - \overline{y}^J\| \subseteq \mathcal{W}^{J\setminus\{j\}}$$

since $\mathcal{W}^{J\setminus\{j\}} \subseteq \mathcal{W}^J$ for all $j \in J$.

As the third (and central) step, we use Farkas' lemma to argue that there exist $\lambda \in \mathbb{R}^J$ with $\lambda_j \geq 0$ for all $j \in J$ such that $X\overline{w}^J = \sum_{j \in J} \lambda_j X \overline{w}^{J\setminus\{j\}}$.

Assume first that $X\hat{w}^J \neq \overline{y}^J$. Then we must have that $\hat{w}^J \in \partial\mathcal{W}^J$. Indeed, if $\hat{w}^J \in \mathcal{W}^J \setminus \partial\mathcal{W}^J$ then there exists some $\varepsilon > 0$ such that $\hat{w}^J(\varepsilon) = \hat{w}^J\,(1-\varepsilon) + \overline{w}^J\,\varepsilon \in \mathcal{W}^J$, for which $\|X\hat{w}^J(\varepsilon) - \overline{y}^J\| = \|X(\hat{w}^J(\varepsilon) - \overline{w}^J)\| = (1-\varepsilon)\|X(\hat{w}^J - \overline{w}^J)\| < \|X\hat{w}^J - \overline{y}^J\|$, contradicting the choice of $\widehat{W}^J$ and $\hat{w}^J$. Hence $\hat{w}^J \in \partial\mathcal{W}^J$, so there is some j with $\hat{w}^J \in \mathcal{W}^{J\setminus\{j\}}$, which implies that $\hat{w}^{J\setminus\{j\}} = \hat{w}^J$ and $X\hat{w}^J = X\hat{w}^{J\setminus\{j\}}$. This means that we can choose λ as the indicator for component j.

Assume now that $X\hat{w}^J = \overline{y}^J$, and that there exists no such λ. Then, by Farkas' lemma, there exists $v \in \mathbb{R}^T \setminus \{\mathbf{0}\}$ such that $v'X\hat{w}^J < 0$ and $v'X\hat{w}^{J\setminus\{j\}} \geq 0$ for all $j \in J$. Define the projection matrix $\Pi = \frac{vv'}{v'v} \in \mathbb{R}^{T\times T}$, and let $\mathcal{W}^* = \arg\min_{w \in \mathcal{W}^J; \Pi X(w - \hat{w}^J) = X(w - \hat{w}^J)} v'Xw \subseteq \mathcal{W}^J$. Then the minimum is attained at a boundary point $w^* \in \mathcal{W}^* \cap \partial\mathcal{W}^J$ of the feasible set. Indeed, the feasible set is non-empty since it includes $\hat{w}^J$, and it is compact and convex. The minimum of the linear function is therefore attained at a boundary point, which is in $\partial\mathcal{W}^J$. As a consequence, $w^* \in \mathcal{W}^{J\setminus\{j\}}$ for some $j \in J$. Since $\hat{w}^J \in \mathcal{W}^J$, we have that $v'Xw^* \leq v'X\hat{w}^J < v'X\hat{w}^{J\setminus\{j\}}$. Hence there is some $\varepsilon \in (0,1]$ such that $\hat{w}^{J\setminus\{j\}}(\varepsilon) = \hat{w}^{J\setminus\{j\}}\,(1-\varepsilon) + w^*\,\varepsilon \in \mathcal{W}^{J\setminus\{j\}}$ fulfills $v'X\hat{w}^{J\setminus\{j\}}(\varepsilon) = v'X\hat{w}^J$. Since we therefore have $\Pi X\hat{w}^{J\setminus\{j\}}(\varepsilon) = \Pi X\hat{w}^J$, as well as $(\mathbb{I} - \Pi)X\hat{w}^{J\setminus\{j\}}(\varepsilon) = (\mathbb{I} - \Pi)X(\hat{w}^{J\setminus\{j\}}\,(1-\varepsilon) + \varepsilon\hat{w}^J)$ since $\Pi X(w^* - \hat{w}^J) = X(w^* - \hat{w}^J)$, we have that

$$\begin{aligned}
&\|X\hat{w}^{J\setminus\{j\}}(\varepsilon) - \overline{y}^J\|^2 = \|X(\hat{w}^{J\setminus\{j\}}(\varepsilon) - \hat{w}^J)\|^2 \\
&= \|\Pi X(\hat{w}^{J\setminus\{j\}}(\varepsilon) - \hat{w}^J)\|^2 + \|(\mathbb{I} - \Pi)X(\hat{w}^{J\setminus\{j\}}(\varepsilon) - \hat{w}^J)\|^2 = 0 + (1-\varepsilon)^2\|(\mathbb{I} - \Pi)X(\hat{w}^{J\setminus\{j\}} - \hat{w}^J)\|^2 \\
&< \|\Pi X(\hat{w}^{J\setminus\{j\}} - \hat{w}^J)\|^2 + \|(\mathbb{I} - \Pi)X(\hat{w}^{J\setminus\{j\}} - \hat{w}^J)\|^2 = \|X(\hat{w}^{J\setminus\{j\}} - \hat{w}^J)\|^2 = \|X\hat{w}^{J\setminus\{j\}} - \overline{y}^J\|^2,
\end{aligned}$$

contradicting the choice of $\hat{w}^{J\setminus\{j\}}$. Hence, such λ must exist.

As the fourth step, we expand the previous result on fitted values to the weights themselves in the case of penalized synthetic control, and show that the weights sum to one in that case. To this end, note that we can write the penalized synthetic control estimator from (2) as $\hat{w}^J_\eta = \arg\min_{w \in \mathcal{W}^J} \|Xw - y\|^2 + \eta\|w\|^2$. Write now $\tilde{X}^J_\eta = (X_J'X_J + \eta\mathbb{I})^{1/2} \in \mathbb{R}^{J\times J}$ for the symmetric positive-definite matrix square root of the symmetric positive-definite $X_J'X_J + \eta\mathbb{I}$, where X_J is a matrix of the columns of X with index in J, and $\tilde{y}^J_\eta = (\tilde{X}^J_\eta)^{-1}X_J'y \in \mathbb{R}^J$. For w_J the entries of $w \in \mathcal{W}^J$ corresponding to the index set J, we find

$$\begin{aligned}
&\|Xw - y\|^2 + \eta\|w\|^2 = \|X_Jw_J - y\|^2 + \eta\|w_J\|^2 \\
&= w_J'X_J'X_Jw_J - 2w_J'X_J'y + y'y + \eta w_J'w_J = w_J'(X_J'X_J + \eta\mathbb{I})w_J - 2w_J'X_J'y + y'y \\
&= w_J'\tilde{X}^{J\prime}_\eta\tilde{X}^J_\eta w_J - 2w_J'\tilde{X}^{J\prime}_\eta\left((\tilde{X}^J_\eta)^{-1}X_J'y\right) + y'y = \|\tilde{X}^J_\eta w_J - \tilde{y}^J_\eta\|^2 - \|\tilde{y}^J_\eta\|^2 + \|y\|^2.
\end{aligned}$$

Hence, we can write (noting that $\mathcal{W}^{J\setminus\{j\}} \subseteq \mathcal{W}^J$)

$$\hat{w}^J_\eta = \underset{w \in \mathcal{W}^J}{\arg\min} \|\tilde{X}^J_\eta w_J - \tilde{y}^J_\eta\|, \qquad w^{J\setminus\{j\}}_\eta = \underset{w \in \mathcal{W}^{J\setminus\{j\}}}{\arg\min} \|\tilde{X}^J_\eta w_J - \tilde{y}^J_\eta\|,$$

so we can interpret penalized synthetic control on units J and $J \setminus \{j\}$ with time periods $\{1, \dots, T\}$ and the original outcomes as non-penalized synthetic control on units J and $J \setminus \{j\}$ with time periods J and transformed outcomes, where we note that the synthetic-control solutions are unique in this case. Hence, we can apply the previous result to conclude that there exists $\lambda_\eta \in \mathbb{R}^J$ with $\lambda_{\eta,j} \geq 0$ for all $j \in J$ such that $\tilde{X}^J_\eta \tilde{w}^J_\eta = \sum_{j \in J} \lambda_{\eta,j} \tilde{X}^J_\eta \tilde{w}^{J \setminus \{j\}}_\eta$. Since $\tilde{X}^J_\eta$ is invertible, it now also follows that $\tilde{w}^J_\eta = \sum_{j \in J} \lambda_{\eta,j} \tilde{w}^{J \setminus \{j\}}_\eta$. Since also $\tilde{w}^J_\eta \in \mathcal{W}$ and $\tilde{w}^{J \setminus \{j\}}_\eta \in \mathcal{W}$ for all $j \in J$, we have that $\sum_{j \in J} \lambda_{\eta,j} = \sum_{j \in J} \lambda_{\eta,j} \mathbf{1}' \tilde{w}^{J \setminus \{j\}}_\eta = \mathbf{1}' \tilde{w}^J_\eta = 1$. This establishes the main claim of the proposition for penalized synthetic control.

As the fifth and final step, we derive the main result on minimum-norm synthetic control from the above results on penalized synthetic control. Consider some sequence $(\eta_\iota)_{\iota=1}^\infty$ in $(0, \infty)$ with $\eta_\iota \to 0$, and for every ι apply the previous step to the penalized synthetic control estimator with penalty η_ι to obtain a weight vector $\lambda_{\eta_\iota} \in \Lambda^J = \left\{\lambda \in [0,1]^J; \sum_{j=1}^J \lambda_j = 1\right\}$. Since Λ^J is compact, $(\lambda_{\eta_\iota})_{\iota=1}^\infty$ must have a converging subsequence with some limit $\lambda \in \Lambda^J$. Using the limit along this subsequence, we have that

$$\hat{w}^J = \lim_{\iota \to \infty} \hat{w}^J_{\eta_\iota} = \lim_{\iota \to \infty} \sum_{j \in J} \lambda_{\eta_\iota, j} \tilde{w}^{J \setminus \{j\}}_{\eta_\iota} = \sum_{j \in J} \left(\lim_{\iota \to \infty} \lambda_{\eta_\iota, j} \right) \left(\lim_{\iota \to \infty} \tilde{w}^{J \setminus \{j\}}_{\eta_\iota} \right) = \sum_{j \in J} \lambda_j \tilde{w}^{J \setminus \{j\}}. \quad \square$$

Proof of Proposition 5. By Jensen's inequality applied to an average over the bounds in (3),

$$\mathrm{E}[(y - \hat{f}^*(x))^2] \leq \frac{1}{|J|!} \sum_\pi \mathrm{E}[(y - \hat{f}^*_\pi(x))^2] \leq \sum_{j \in J} \mathrm{E}\left[\underbrace{\frac{1}{|J|!} \sum_\pi \hat{\lambda}_{\pi(j)}}_{= \frac{1}{|J|}} (y - \hat{f}^j(x))^2 \right]. \quad \square$$

C Details of the Empirical Illustrations

C.1 Many-Regressor Linear Least-Squares on CPS Data

We utilize the publicly available[2] CPS control and NSW experimental control datasets, drawn from the study presented in LaLonde (1986) as used by Dehejia and Wahba (1999, 2002). The resulting data has 15,992 observations for CPS and 260 for NSW, with both datasets containing an identical set of variables, detailed in Table 1.

Variable	Data Type	Description
`age`	Discrete	Age
`education`	Discrete	Years of education
`black`	Dummy	Black
`hispanic`	Dummy	Hispanic
`married`	Dummy	Marital status
`nodegree`	Dummy	Lack of college degree
`re74`	Continuous	Income in 1974
`re75`	Continuous	Income in 1975
`re78`	Continuous	Income in 1978

Table 1: CPS and NSW dataset variables

We use `re78` as the outcome variable and all other variables as covariates. In order to achieve high dimensionality, we first discretize the continuous income covariates into 50 bins via quantile binning. We then construct a series of dummies for each discrete variable, corresponding to indicators for each discretized value. We then interact all these dummy variables, as well as those covariates which were originally dummies, taking care not to interact those which are mutually exclusive (e.g. originating

[2] `users.nber.org/~rdehejia/data/.nswdata2.html`. We use the files corresponding to `cps_controls.txt` and `nswre74_control.txt`.

from the same original covariate or corresponding to race). We then drop any interactions that are zero for all observations in the data. The resulting transformed dataset contains 8,408 dummy covariates, as well as the unmodified outcome variable. In order to ensure that the covariate matrix is full rank for an arbitrary subset of columns, we go on to add iid $\mathcal{N}(0, 0.0004)$ noise to each of the covariate values (again leaving the outcome variable unaffected). We then select a random subset of 3,000 observations from the CPS dataset as our in-sample set, keeping the remaining 12,992 CPS and 260 NSW observations as out-of-sample sets.

For fitting models of varying complexity, we randomly permute the order of the columns of the covariate matrix; denote the resulting matrix as X. We then add an intercept and iterate over varying levels of complexity ℓ, ranging from 1 to 8,409, corresponding to the number of covariates that we will use for estimation. We then estimate the OLS coefficient vector $\hat{\beta}^\ell = X^{\ell\dagger} y$, where † denotes the Moore–Penrose pseudoinverse, and calculate the out-of-sample prediction error

$$\mathrm{RMSE}(\ell) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i^* - x_i^{*\top}\hat{\beta}^\ell)^2}$$

Where X^*, y^* denote the out-of-sample covariate and outcome variables, respectively, and can correspond to either the CPS or NSW held-out samples. We also calculate in-sample prediction error analogously, using the in-sample covariates and outcomes X, y. In order to smooth out the effects of the random ordering of columns, we repeat this exercise for five different random orderings and take a pointwise average to obtain smooth RMSE vs. complexity and coefficient vector norm vs. complexity curves (Figures 1 and 7, respectively).

C.2 Many-Unit Synthetic Control on Smoking Data

For our synthetic-control exercise, we utilize public data from the Centers for Disease Control and Prevention[3] containing annual cost, revenue, tax, and quantity data for cigarette sales by state for the years 1970 to 2019. We follow the approach of Abadie et al. (2010) in using synthetic control to estimate per-capita cigarette pack consumption for the target state, California, as a function of the other 49 states and Washington, D.C. For our evaluation, we utilize two years of data (1987 and 1988) as a hold-out sample and fit the model on three years (1984 to 1986). All of our data precedes the year in which anti-smoking legislation took effect in California (1989).

We begin by selecting a random subset of 20 states to serve as our donor pool, for computational tractability. We then select a level of complexity ℓ and select a subset of ℓ states from the chosen 20. Using that subset, we then estimate synthetic control weights based on the in-sample period, choosing convex weight vector $\hat{w}^\ell \in \mathcal{W} = \{w \in [0,1]^N; \sum_{i=1}^{n} w_i = 1\}$ as described in Section 3.1:

$$\hat{w}^\ell = \operatorname*{arg\,min}_{w \in \widehat{\mathcal{W}}^\ell} \|w\| \qquad \widehat{\mathcal{W}}^\ell = \operatorname*{arg\,min}_{w \in \mathcal{W}; w_j = 0 \forall j \notin J} \sum_{t=1984}^{1986} (y_{0t} - \sum_{i=1}^{\ell} w_i y_{it})^2.$$

Here, y_0 denotes the target state, California. We then compute the out-of-sample prediction error as:

$$\mathrm{RMSE}(\ell) = \sqrt{\frac{1}{2}\sum_{t=1987}^{1988}(y_{0t} - \sum_{i=1}^{\ell}\hat{w}_i^\ell y_{it})^2}$$

We then iterate over all $\binom{20}{\ell}$ possible combinations of donor units for the given complexity level and take the average RMSE value to be the predictive error for the given complexity level. We vary ℓ from 1 to 20 to trace out the curve of synthetic control prediction risk vs. complexity (Figure 4).

[3] chronicdata.cdc.gov/Policy/The-Tax-Burden-on-Tobacco-1970-2019/7nwe-3aj9.