Forward selection and post-selection inference in factorial designs

Lei Shi

UC Berkeley, Biostatistics

October 7, 2024

joint work with Peng Ding (Berkeley Statistics) and Jingshen Wang (Berkeley Biostatistics)

https://arxiv.org/abs/2301.12045

Outline

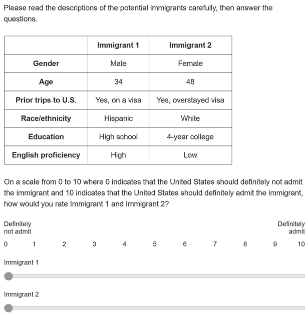
Outline

Motivation

- Factorial experiments: a powerful design to accommodate multiple factors and offer opportunities to estimate both the main causal effects of factors and their interactions;
- Quick example: a combination of drugs, a combination of agricultural conditions, . . .
- We focus on 2^K replicated full factorial experiments, with K binary factors and at least two units $N(z) \geq 2$ within arm z.
- Classical regimes: small K and a large number of replications;
- should we go beyond? can we go beyond?

Motivation

- Conjoint survey experiments: modern factorial experiments, popular in political science;
- **Surveys** that contain **hypothetical profiles** based on **randomized combinations of factors (attributes)**;

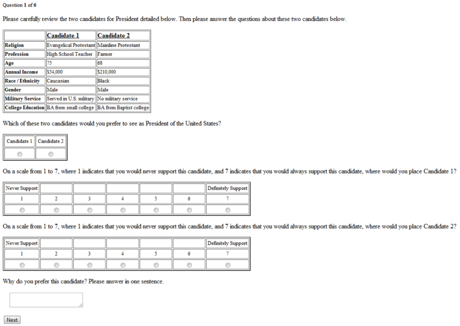


Please read the descriptions of the potential immigrants carefully, then answer the questions.

	Immigrant 1	Immigrant 2
Gender	Male	Female
Age	34	48
Prior trips to U.S.	Yes, on a visa	Yes, overstayed visa
Race/ethnicity	Hispanic	White
Education	High school	4-year college
English proficiency	High	Low

On a scale from 0 to 10 where 0 indicates that the United States should definitely not admit the immigrant and 10 indicates that the United States should definitely admit the immigrant, how would you rate Immigrant 1 and Immigrant 2?

Figure: Survey on Immigrants



Question 1 of 6

Please carefully review the two candidates for President detailed below. Then please answer the questions about these two candidates below.

	Candidate 1	Candidate 2
Religion	Evangelical Protestant	Mainline Protestant
Profession	High School Teacher	Farmer
Age	75	68
Annual Income	$54,000	$210,000
Race / Ethnicity	Caucasian	Black
Gender	Male	Male
Military Service	Served in U.S. military	No military service
College Education	BA from small college	BA from Baptist college

Which of these two candidates would you prefer to see as President of the United States?

On a scale from 1 to 7, where 1 indicates that you would never support this candidate, and 7 indicates that you would always support this candidate, where would you place Candidate 1?

On a scale from 1 to 7, where 1 indicates that you would never support this candidate, and 7 indicates that you would always support this candidate, where would you place Candidate 2?

Why do you prefer this candidate? Please answer in one sentence.

Figure: Survey on Presidential Candidates

Motivation

- Can handle large K and N, powered by computers and web-based technology
- Some numbers from conjoint survey experiments:

Experiment	**Reference**	K	Q	N	N_0
Immigrant admission experiment	[Zhirkov, 2022]	6	$2^6 = 64$	$\sim 28,000$	~ 430
U.S. presidential election	[Caughey et al., 2019]	12	$2^{12} = 4096$	$\sim 30,000$	~ 8
Aluminum packaging characteristics	[Li et al., 2013]	7	$2^6 * 4 = 256$	$\sim 15,000$	~ 60

Note: K is the number of factors, Q is the number of treatment combinations, N it the number of units (hypothetical profiles), N_0 is the average replications per arm.

Motivation

Challenge	**Goal**
Too many causal parameters in large K settings: $2^K - 1$ factorial effects	Identify a few significant factorial effects and rule out negligible effects
Need principled methods for effect selection in a design-based framework	Develop a procedure that fully respects the principles of factorial experiments
Few discussions on how effect selection affects estimation and inference	Exploring post-selection inference for general causal parameters

Table: Challenge & Goal

Outline

1. Motivation
2. A Tutorial on Randomized Factorial Experiments
3. Forward Factorial Selection
4. Inference on General Causal Estimands under Consistent Model Selection
5. Inference on General Causal Estimands under Inconsistent Model Selection
6. Case Study: Preference for U.S. Presidential Candidates
7. Concluding Remarks

2^K Factorial Experiments

- K factors: $z_k \in \{-1, 1\}$ for $k = 1, \ldots, K$;
- $Q = 2^K$ treatment combinations:

$$\mathcal{T} = \{z = (z_1, \ldots, z_K) \mid z_k \in \{-1, 1\} \text{ for } k = 1, \cdots, K\} \quad \text{with} \quad |\mathcal{T}| = Q.$$

- N: sample size; $N(z)$: sample size under treatment $z \in \mathcal{T}$;
- $Y_i(z)$: potential outcome if i-th unit was assigned to the treatment z;
- Z_i: the observed treatment for unit i;
- $Y_i = Y_i(Z_i)$: the observed outcome for unit i;
- **Complete randomization**: the treatment vector $(Z_1, \ldots, Z_N)$ is a random permutation of a vector with prespecified number $N(z)$ of the treatment combination z, for all $z \in \mathcal{T}$.

Factorial effects: main effects

- Defined by a set of contrast vectors [Wu and Hamada, 2011, Dasgupta et al., 2015];
- Define population mean vector $\overline{Y} = (\overline{Y}(z))_{z\in\mathcal{T}}$, with elements $\overline{Y}(z) = N^{-1}\sum_{i=1}^{N} Y_i(z)$.
- Contrast vectors for main effects:

$$g_{\{k\}} = \{g_{\{k\}}(z)\}_{z\in\mathcal{T}}, \text{ where } g_{\{k\}}(z) = z_k.$$

  (Vectorize treatment indicators for each factor)
- Main effects:

$$\tau_{\{k\}} = Q^{-1} \cdot g_{\{k\}}^\top \overline{Y} \quad \text{for} \quad k \in [K]. \tag{1}$$

Factorial effects: interaction effects

- Contrast vectors for interaction effects: for $\mathcal{K} = \{k_1, \cdots, k_d\}$ with $d \geq 2$, do element-wise product of $g_{\{k_1\}}, \ldots, g_{\{k_d\}}$:

$$g_{\mathcal{K}} = \{g_{\mathcal{K}}(z)\}_{z \in \mathcal{T}}, \text{ where } g_{\mathcal{K}}(z) = \prod_{k \in \mathcal{K}} g_{\{k\}}(z) = \prod_{k \in \mathcal{K}} z_k. \tag{2}$$

- The interaction effect is:

$$\tau_{\mathcal{K}} = Q^{-1} \cdot g_{\mathcal{K}}^\top \overline{Y} \quad \text{for} \quad \mathcal{K} \subset [K]. \tag{3}$$

- $\tau_{\mathcal{K}}$ is a parent effect of $\tau_{\mathcal{K}'}$ if $\mathcal{K} \subset \mathcal{K}'$ and $|\mathcal{K}| = |\mathcal{K}'| - 1$.
- Contrast matrix: stack the $g_{\mathcal{K}}$'s into a $Q \times Q$ orthogonal matrix:

$$G = (g_\varnothing, g_{\{1\}}, \ldots, g_{\{K\}}, g_{\{1,2\}}, \ldots, g_{\{K-1,K\}}, \ldots, g_{[K]}). \tag{4}$$

Estimation with Factor-based Regression

- Factor-based regression is a model-assisted strategy for estimating factorial effects.
  - Saturated regression: $Y_i \sim t_i$;
  - Unsaturated regression: $Y_i \sim$ a subvector of t_i;
    - Denote the subvector of t_i as $t_{i,\mathbb{M}}$ for a collection of factor combinations $\mathbb{M}$;
    - We refer to $\mathbb{M}$ as a working model;
  - $\widehat{\tau}(\mathbb{M})$: estimated coefficients with working model $\mathbb{M}$;
  - $\tau(\mathbb{M})$: the collection of true factorial effects in $\mathbb{M}$;
  - With weighted least squares: $\widehat{\tau}(\mathbb{M})$ are unbiased for $\tau(\mathbb{M})$ [Zhao and Ding, 2021].

A Running Example of 2^3 Factorial Experiment

- Three binary factors z_1, z_2, and z_3 with 8 treatment combinations:

$$\mathcal{T} = \{(---),(--+),(-+-),(-++),(+--),(+-+),(++-),(+++)\}.$$

- Factorial effects $\tau = \frac{1}{2^3} G^\top \overline{Y} \triangleq \left(\tau_\varnothing, \tau_{\{1\}}, \tau_{\{2\}}, \tau_{\{3\}}, \tau_{\{1,2\}}, \tau_{\{1,3\}}, \tau_{\{2,3\}}, \tau_{\{1,2,3\}}\right)^\top$, where

$$G = \begin{array}{c} \\ (---) \\ (--+) \\ (-+-) \\ (-++) \\ (+--) \\ (+-+) \\ (++-) \\ (+++) \end{array} \begin{array}{c} \begin{array}{cccccccc} \tau_\varnothing & \tau_{\{1\}} & \tau_{\{2\}} & \tau_{\{3\}} & \tau_{\{1,2\}} & \tau_{\{1,3\}} & \tau_{\{2,3\}} & \tau_{\{1,2,3\}} \end{array} \\ \begin{pmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \end{array}.$$

A Running Example of 2^3 Factorial Experiment

- Saturated regression: $Y_i \sim t_i$ with

$$g_i = \Big[1,\ g_{\{1\}}(Z_i),\ g_{\{2\}}(Z_i),\ g_{\{3\}}(Z_i),\ g_{\{2,3\}}(Z_i),\ g_{\{1,3\}}(Z_i),\ g_{\{1,2\}}(Z_i),\ g_{\{1,2,3\}}(Z_i)\Big].$$

- Unsaturated regression: $\mathbb{M} = \{\varnothing, \{1\}, \{1,2\}, \{1,3\}, \{1,2,3\}\}$, $Y_i \sim t_{\mathbb{M},i}$ with

$$g_{i,\mathbb{M}} = \Big[1,\ g_{\{1\}}(Z_i),\ g_{\{1,3\}}(Z_i),\ g_{\{1,2\}}(Z_i),\ g_{\{1,2,3\}}(Z_i)\Big]$$

and the weight for unit i equals $1/N_i = 1/N(Z_i)$.

Motivation

Challenge	Goal
Too many causal parameters in large K settings: $2^K - 1$ factorial effects	Identify a few significant factorial effects and rule out negligible effects
Need **principled** methods for effect selection in a design-based framework	Develop a procedure that fully respects the principles of factorial experiments
Few discussions on how effect selection affects estimation and inference	Exploring post-selection inference for general causal parameters

Table: Challenge & Goal

Outline

Forward Selection: the High Level Ideas

- The fundamental challenge: too many causal parameters
- Three principles summarized by [Wu and Hamada, 2011]:
  - **Effect Hierarchy Principle.** (i) Lower-order effects are more likely to be important than higher-order effects. (ii) Effects of the same order are equally likely to be important.
  - **Effect Sparsity Principle.** The number of relatively important effects in a factorial experiment is small.
  - **Effect Heredity Principle.** For an interaction to be significant, all of its parent effects should be significant (strong heredity) or at least one of its parent effects should be significant (weak heredity).
- Motivates a natural selection procedure that proceeds in a forward style!

Forward Selection: A Running Example

- Consider a 2^3 factorial experiment, and we do forward selection with Bonferroni corrected marginal t-tests;

Step	Working Model
1. Start with the intercept	$\widehat{\mathbb{M}} = \{\varnothing\}$
2. Add all the main effects	$\widehat{\mathbb{M}} = \{\varnothing, \{1\}, \{2\}, \{3\}\}$
3. Regress $Y_i \sim g_{i,\widehat{\mathbb{M}}}$ with weights $1/N_i$	$\widehat{\mathbb{M}} = \{\varnothing, \{1\}, \{2\}, \{3\}\}$
4. Do t-tests on main effects with level $\alpha_1/\lvert\widehat{\mathbb{M}}\backslash\{\varnothing\}\rvert$ and drop the non-rejections	$\widehat{\mathbb{M}} = \{\varnothing, \{1\}, \{2\}, \{1,2\}\}$
5. Add two way interactions under strong heredity	$\widehat{\mathbb{M}} = \{\varnothing, \{1\}, \{2\}, \{1,2\}\}$
6. Regress $Y_i \sim g_{i,\widehat{\mathbb{M}}}$ with weights $1/N_i$	$\widehat{\mathbb{M}} = \{\varnothing, \{1\}, \{2\}, \{1,2\}\}$
7. Do t-tests on main effects with level $\alpha_2/\lvert\widehat{\mathbb{M}}\backslash\{\varnothing, \{1\}, \{2\}\}\rvert$ and drop the non-rejections	$\widehat{\mathbb{M}} = \{\varnothing, \{1\}, \{2\}\}$
8. No two-way effects identified; return $\widehat{\mathbb{M}}$	$\widehat{\mathbb{M}} = \{\varnothing, \{1\}, \{2\}\}$

Forward Selection: A Dissection

- The selection procedure is iterated in a forward style:

$$\widehat{\mathbb{M}}_1 \xrightarrow{\texttt{H}} \widehat{\mathbb{M}}_{2,+} \xrightarrow{\widehat{\texttt{S}}} \widehat{\mathbb{M}}_2 \cdots \xrightarrow{\widehat{\texttt{S}}} \widehat{\mathbb{M}}_{d-1} \xrightarrow{\texttt{H}} \widehat{\mathbb{M}}_{d,+} \xrightarrow{\widehat{\texttt{S}}} \widehat{\mathbb{M}}_d \rightarrow \cdots \xrightarrow{\widehat{\texttt{S}}} \widehat{\mathbb{M}}_D,$$

  - Respects the *Effect Hierarchy Principle*: forward-type algorithm
  - Respects the *Effect Sparsity Principle*: S-step
  - Respects the *Effect Heredity Principle*: H-step
- Generates a highly interpretable working model
- Compatible with many selection methods: Marginal t-test [Wasserman and Roeder, 2009], Lasso [Zhao and Yu, 2006], SIS [Fan and Lv, 2008]...

Forward Selection: Selection Consistency

- Forward procedure gives a consistent model selection result under the design-based framework:

Theorem (Consistent selection property)

Under some regularity conditions,

$$\lim_{N\to\infty} \mathbb{P}\left\{\widehat{\mathbb{M}} = \mathbb{M}^\star_{1:D}\right\} = 1.$$

- Relies on some novel permutational Berry-Esseen bounds: [Shi and Ding, 2022].

Outline

Causal Estimands and Estimators

- Estimands: a linear combination of average potential outcomes.

$$\gamma = \sum_{z \in \mathcal{T}} f(z)\overline{Y}(z) \triangleq f^\top \overline{Y}, \tag{5}$$

- Examples:
  - factorial effects: $f = g_{\{k\}}$;
  - ATE over two levels: $f = (1, -1, 0, \ldots, 0)^\top$.

Estimators

- Without factor selection: plug-in estimator

$$\widehat{\gamma} = f^\top \widehat{Y} = \sum_{z \in \mathcal{T}} f(z)\widehat{Y}(z), \quad \widehat{v}^2 = f^\top \widehat{V}_{\widehat{Y}} f = \sum_{z \in \mathcal{T}} f(z)^2 N(z)^{-1} \widehat{S}(z,z). \tag{6}$$

- With factor selection: Restricted least squares. let $f[\mathbb{M}] = Q^{-1}G(\cdot,\mathbb{M})G(\cdot,\mathbb{M})^\top f$, and

$$\widehat{\gamma}_{\mathrm{R}} = f[\widehat{\mathbb{M}}]^\top \widehat{Y} \quad \text{and} \quad \widehat{v}^2_{\mathrm{R}} = f[\widehat{\mathbb{M}}]^\top \widehat{V}_{\widehat{Y}} f[\widehat{\mathbb{M}}]. \tag{7}$$

Statistical property

- Theoretical property: under some regularity condition,
  - Asymptotic normality: $(\widehat{\gamma}_{\mathrm{R}} - \gamma)/v_{\mathrm{R}} \rightsquigarrow \mathcal{N}(0,1)$, where $v_{\mathrm{R}}^2 = f^{\star\top} V_{\widehat{Y}} f^{\star}$.
  - Variance convergence: $N(\widehat{v}_{\mathrm{R}}^2 - v_{\mathrm{R,lim}}^2) \xrightarrow{\mathrm{P}} 0, \quad v_{\mathrm{R,lim}}^2 \geq v_{\mathrm{R}}^2$.
- For inference on factorial effects, the behavior is similar;
- For inference on ATE,
  - Sparse f and small N_0
  - $\widehat{\gamma}_{\mathrm{R}}$ converges in distribution but $\widehat{\gamma}$ fails
  - v_{R}^2 has efficiency gain with large Q suggested by

$$\frac{v_{\mathrm{R}}^2}{v^2} \leq \kappa(V_{\widehat{Y}}) \cdot \frac{s^{\star}|\mathbb{M}^{\star}|}{Q}.$$

Outline

Inconsistent Selection

- Consistent selection might fail when high-order effects are small.
- Drawback of post-selection inference methods in factorial designs:
  - Data splitting: relies on independence and exchangeability; inference on a random parameter
  - Selective inference: relies on specific selection methods and modeling assumptions

Strategy in the presence of inconsistent Selection

- Under Selection: Select factorial effects up to level $d < K$ and exclude effects beyond level d.
- When to apply:
  - The number of active lower-order interactions is large
  - High-order interactions are nuisance parameters or not of interest
  - Domain knowledge indicates that higher-order interactions are negligible
- Model selection is consistent for lower-order effects

Strategy in the presence of inconsistent Selection

- Over Selection: Select factorial effects up to some level $d < K$ and select higher-order effects by the heredity principle.
- When to apply:
  - Non-negligible higher-order interactions exist
  - The research question targets a more general parameter beyond factorial effects
- Final working model over-selects the truth

Outline

Conjoint survey experiment regarding U.S. presidential candidates

- a conjoint survey experiment regarding U.S. citizens' preferences across presidential candidates studied by [Hainmueller et al., 2014]
- we include 7 attributes (factors) of the imaginary candidate profiles: military service (z_1), religion (z_2), college education (z_3), annual income (z_4), racial/ethnic background (z_5), age (z_6), gender (z_7)
- $K = 7$ factors (with $Q = 2^7 = 128$ treatment combinations) and $N = 3456$ profiles. Each treatment combination contains 27 respondents.

Conjoint survey experiment regarding U.S. presidential candidates

Table: Model Selection Results for the Presidential Candidate Experiment

Selection Strategy	Selected Working Model
Forward + Strong Heredity	τ_2, τ_3, τ_4, τ_6, τ_7, τ_{23}, τ_{36}, τ_{46}, τ_{47}, τ_{67}, τ_{467}
Forward + Weak Heredity	τ_2, τ_3, τ_4, τ_6, τ_7, τ_{12}, τ_{23}, τ_{13}, τ_{35}, τ_{36}, τ_{14}, τ_{46}, τ_{47}, τ_{56}, τ_{67}, τ_{57}
Forward + No Heredity	τ_2, τ_3, τ_4, τ_6, τ_7, τ_{12}, τ_{23}, τ_{13}, τ_{35}, τ_{36}, τ_{14}, τ_{46}, τ_{47}, τ_{56}, τ_{67}, τ_{57}
No Forward	τ_3, τ_6, τ_{14}

- (Row 1 & 4) Forward selection leads to more terms in the working model, while the full LASSO procedure selects an overly sparse one
- (Row 1 & 2) Weak and strong heredity produces parsimonious and interpretable working models
- (Row 2 & 3) Weak and no heredity coincide, validating the plausibility of heredity

Outline

Concluding Remarks

- Further directions for exploration:
  - extension of the theory to general factorial designs with multi-valued factors under more complicated notations
  - covariate adjustment in factorial experiments; [Lin, 2013, Zhao and Ding, 2023]
  - extension of the framework to observational studies by incorporating propensity score and outcome model estimation
- Acknowledgement: Special thanks for the acknowledgment of the SFASA!

Reference I

Caughey, D., Katsumata, H., and Yamamoto, T. (2019).
Item response theory for conjoint survey experiments.
Technical report, Working Paper.

Dasgupta, T., Pillai, N. S., and Rubin, D. B. (2015).
Causal inference from 2^K factorial designs by using potential outcomes.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77:727–753.

Fan, J. and Lv, J. (2008).
Sure independence screening for ultrahigh dimensional feature space.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70(5):849–911.

Hainmueller, J., Hopkins, D. J., and Yamamoto, T. (2014).
Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments.
Political Analysis, 22(1):1–30.

Reference II

Li, W., Nachtsheim, C. J., Wang, K., Reul, R., and Albrecht, M. (2013).
Conjoint analysis and discrete choice experiments for quality improvement.
Journal of Quality Technology, 45(1):74–99.

Lin, W. (2013).
Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique.
Annals of Applied Statistics, 7(1):295–318.

Shi, L. and Ding, P. (2022).
Berry-esseen bounds for design-based causal inference with possibly diverging treatment levels and varying group sizes.
arXiv preprint arXiv:2209.12345.

Wasserman, L. and Roeder, K. (2009).
High dimensional variable selection.
Annals of Statistics, 37(5A):2178.

Reference III

Wu, C. J. and Hamada, M. S. (2011).
Experiments: Planning, Analysis, and Optimization.
Hoboken, NJ: John Wiley & Sons.

Zhao, A. and Ding, P. (2021).
Covariate-adjusted fisher randomization tests for the average treatment effect.
Journal of Econometrics, 225(2):278–294.

Zhao, A. and Ding, P. (2023).
Covariate adjustment in multiarmed, possibly factorial experiments.
Journal of the Royal Statistical Society Series B: Statistical Methodology, 85(1):1–23.

Zhao, P. and Yu, B. (2006).
On model selection consistency of lasso.
The Journal of Machine Learning Research, 7:2541–2563.

Reference IV

Zhirkov, K. (2022).
Estimating and using individual marginal component effects from conjoint experiments.
Political Analysis, 30(2):236–249.