Biometrika (2021), **00**, 0, *pp.* 1–17
Printed in Great Britain
doi: 10.1093/biomet/asab051

Regression-based causal inference with factorial experiments: estimands, model specifications and design-based properties

BY ANQI ZHAO

Department of Statistics and Data Science, National University of Singapore, 117546 Singapore
staza@nus.edu.sg

AND PENG DING

Department of Statistics, University of California, Berkeley,
425 Evans Hall, Berkeley, California 94720, U.S.A.
pengdingpku@berkeley.edu

SUMMARY

Factorial designs are widely used because of their ability to accommodate multiple factors simultaneously. Factor-based regression with main effects and some interactions is the dominant strategy for downstream analysis, delivering point estimators and standard errors simultaneously via one least-squares fit. Justification of these convenient estimators from the design-based perspective requires quantifying their sampling properties under the assignment mechanism while conditioning on the potential outcomes. To this end, we derive the sampling properties of the regression estimators under a wide range of specifications, and establish the appropriateness of the corresponding robust standard errors for Wald-type inference. The results help to clarify the causal interpretation of the coefficients in these factor-based regressions, and motivate the definition of general factorial effects to unify the definitions of factorial effects in various fields. We also quantify the bias-variance trade-off between the saturated and unsaturated regressions from the design-based perspective.

Some key words: Factorial effect; Potential outcome; Randomization inference; Robust standard error.

1. INTRODUCTION

Factorial designs have become increasingly popular in field experiments in the social sciences (e.g., Duflo et al., 2007; Dasgupta et al., 2015; Branson et al., 2016; Egami & Imai, 2019) as well as in traditional agricultural, industrial and biomedical applications (e.g., Wu & Hamada, 2009). Factor-based regression remains the dominant strategy for downstream analysis (e.g., Karlan & List, 2007; Eriksson & Rooth, 2014; Torres et al., 2021), enabling not only direct estimation of the factorial effects as regression coefficients, but also flexible unsaturated specifications to reduce model complexity. A formal justification of its role in causal inference, however, requires both clearly defining the estimands of interest and deriving the sampling properties of the resulting estimators under the potential outcomes framework.

This article makes several contributions. First, we clarify the causal interpretation of the coefficients in factor-based linear regressions, and propose a location-shift strategy for reproducing the

design-based inference of various factorial effects via least squares. Importantly, we show that the robust covariance, also known as the Eicker–Huber–White covariance, affords an asymptotically conservative estimator of the true sampling covariance from the design-based perspective, justifying its use for large-sample Wald-type inference. Second, we review and clarify the definitions of factorial effects in the causal inference, experimental design, epidemiology and social sciences literature, and extend them to allow for arbitrary weighting schemes to accommodate external validity concerns. Third, we derive for the first time the design-based properties of estimators from unsaturated factor-based regressions, and quantify the bias-variance trade-off between the saturated and unsaturated regressions from the design-based perspective.

We use $Y_i \sim x_i$ to denote the least-squares regression of Y_i on x_i and focus on, not only the causal interpretation of the regression coefficients for estimating the general factorial effects, but also the design-based properties of the robust covariance, for large-sample Wald-type inference. The terms regression, coefficients and robust covariance refer to the numeric outputs of least squares free of any modelling assumptions; we evaluate their sampling properties from the design-based perspective. We omit discussion of the ordinary covariance derived under homoskedasticity owing to its lack of design-based guarantees even with the simple treatment-control experiment (Freedman, 2008).

Let 1_N and 1_Q denote the $N \times 1$ and $Q \times 1$ vectors of ones, respectively. Let $\mathcal{I}(\cdot)$ be the indicator function. Let $[m] = \{1, \ldots, m\}$ be the set of positive integers from 1 to m. For two symmetric matrices M_1 and M_2, write $M_1 \geqslant 0$ if M_1 is positive semidefinite and write $M_1 \leqslant M_2$ or $M_1 \geqslant M_2$ if $M_2 - M_1$ is positive or negative semidefinite, respectively.

2. FRAMEWORK, CAUSAL EFFECTS AND TREATMENT-BASED REGRESSION

Consider an experiment with N units indexed by $i = 1, \ldots, N$ and Q treatment levels indexed by $z \in \mathcal{T} = \{1, \ldots, Q\}$. Let $Y_i(z)$ be the potential outcome of unit i if assigned to level z, and let $\bar{Y}(z) = N^{-1} \sum_{i=1}^{N} Y_i(z)$ be the finite-population average, vectorized as $\bar{Y} = \{\bar{Y}(1), \ldots, \bar{Y}(Q)\}^{\mathrm{T}}$. Let $S = \{S(z, z')\}_{z,z' \in \mathcal{T}}$ be the finite-population covariance matrix of the potential outcomes with $S(z, z') = (N-1)^{-1} \sum_{i=1}^{N} \{Y_i(z) - \bar{Y}(z)\}\{Y_i(z') - \bar{Y}(z')\}$. The goal is to estimate $\tau = G\bar{Y}$ for some contrast matrix G with rows orthogonal to 1_Q. Complete randomization assigns completely at random $N_z \geqslant 2$ units to level z with $\sum_{z \in \mathcal{T}} N_z = N$ and $e_z = N_z/N$. For unit i, let $Z_i \in \mathcal{T}$ denote the treatment level and $Y_i = \sum_{z \in \mathcal{T}} \mathcal{I}(Z_i = z) Y_i(z)$ the observed outcome. Let $\hat{Y}(z) = N_z^{-1} \sum_{i:Z_i=z} Y_i$ be the average observed outcome under treatment level z, vectorized as $\hat{Y} = \{\hat{Y}(1), \ldots, \hat{Y}(Q)\}^{\mathrm{T}}$. Then $\hat{\tau} = G\hat{Y}$ is an intuitive choice for estimating τ.

Design-based inference, also known as randomization inference, concerns the sampling properties of estimators over the distribution of the treatment indicators, conditioning on the potential outcomes (Neyman 1923; Imbens and Rubin 2015). In this paper we focus on complete randomization and assume the following condition for asymptotic properties (Li & Ding, 2017).

Condition 1. As N goes to infinity, for all $z \in \mathcal{T}$ we have that (i) $N_z \geqslant 2$ and e_z has a limit between (0, 1); (ii) $\bar{Y}$ and S have finite limits; and (iii) $\max_{1 \leqslant i \leqslant N} \{Y_i(z) - \bar{Y}(z)\}^2/N \to 0$.

Under complete randomization, $\hat{Y}$ is unbiased for $\bar{Y}$ with covariance $\mathrm{cov}(\hat{Y}) = \mathrm{diag}\{S(z,z)/N_z\}_{z \in \mathcal{T}} - N^{-1}S$. Define $\hat{V} = \mathrm{diag}\{\hat{S}(z,z)/N_z\}_{z \in \mathcal{T}}$, where $\hat{S}(z,z) = (N_z - 1)^{-1} \sum_{i:Z_i=z} \{Y_i - \hat{Y}(z)\}^2$, as a moment estimator of $\mathrm{cov}(\hat{Y})$. It is conservative in the sense that $E(\hat{V}) - \mathrm{cov}(\hat{Y}) = N^{-1}S \geqslant 0$. Condition 1 further ensures that $\hat{Y}$ is asymptotically normal with $N\{\hat{V} - \mathrm{cov}(\hat{Y})\} = S + o_{\mathrm{p}}(1)$ (Li & Ding, 2017). The Wald-type inference of τ can thus be

conducted using $\hat{\tau} = G\hat{Y}$ and $\widehat{\mathrm{cov}}(\hat{\tau}) = G\hat{V}G^{\mathrm{T}}$ as the point estimator and estimated covariance, respectively. It is in general conservative because of the overestimation of the covariance; one exception is when the treatment effects are constant across all units, as specified in the following.

Condition 2. For all $z, z' \in \mathcal{T}$, the $Y_i(z) - Y_i(z') = c(z, z')$ are constant across $i = 1, \ldots, N$. This ensures that the $S(z, z')$ are identical for all $z, z' \in \mathcal{T}$, denoted by $S(z, z') = s_0$.

Treatment-based regression is a convenient tool for computing $\hat{Y}$ and $\hat{V}$ from least squares. The regression $Y_i \sim \mathcal{I}(Z_i = 1) + \cdots + \mathcal{I}(Z_i = Q)$ without an intercept yields a coefficient vector $\hat{\beta}$ and a robust covariance $\hat{V}_0$ that satisfy $\hat{\beta} = \hat{Y}$ and $\hat{V}_0 = \mathrm{diag}(1 - N_z^{-1})_{z\in\mathcal{T}}\hat{V} = \hat{V} + o_{\mathrm{p}}(1)$ (Wu & Ding, 2021, § 3.3). The Wald-type inference of τ can therefore also be conducted using $G\hat{\beta}$ and $G\hat{V}_0G^{\mathrm{T}}$ as the point estimator and estimated covariance, respectively.

This set-up encompasses as a special case the $Q_1 \times \cdots \times Q_K$ factorial experiment, which involves $Q = \prod_{k=1}^{K} Q_k$ treatment levels as the combinations of $K \geqslant 2$ factors with Q_k ($k = 1, \ldots, K$) levels. Treatment-based regression accordingly provides a principled way of studying general factorial experiments. It is nevertheless not the dominant strategy in practice when the estimands of interest take certain special forms. When the goal is to estimate the main effects or interactions of the factors under study, a more prevalent practice is to regress the outcome on the factors themselves, and interpret the coefficients as the corresponding factorial effects of interest. This seemingly straightforward approach has several variants used in different fields, which turn out to target factorial effects under distinct weighting schemes. The first contribution of this work is to unify these variants within a class of location-shifted factor-based regressions, and establish the design-based properties of the resulting coefficients and robust covariances.

More importantly, treatment-based regression is saturated and requires the estimation of $Q = \prod_{k=1}^{K} Q_k \geqslant 2^K$ parameters. This can be demanding in terms of sample size even with a moderate number of factors. Factor-based regression, on the other hand, enables the use of flexible unsaturated specifications that include only the main effects and possibly some lower-order interactions corresponding to the factorial effects of interest. Despite the intuitiveness of such an approach and its dominance in practice, the existing literature on the design-based properties of factor-based regression focuses on saturated specifications (Dasgupta et al., 2015; Lu, 2016), and the theory of their unsaturated counterparts remains an open question. Our second contribution is to fill this gap and establish the design-based properties of unsaturated factor-based regressions.

Because of the notational burden involved in the general setting, we start with the 2^2 and 2^3 experiments to illustrate the main ideas and then unify the results under the 2^K experiment. The results convey all key points for the theory of the general $Q_1 \times \cdots \times Q_K$ experiment. We present the formal theory of the general case in the Supplementary Material.

3. The 2^2 factorial experiment

3.1. *A review of existing strategies*

The 2^2 factorial experiment is the simplest factorial experiment with two binary factors, which we denote by A and B. The $Q = 2^2 = 4$ treatment combinations consist of $\mathcal{T} = \{(00), (01), (10), (11)\}$, indexed by $z = (ab)$ for $a, b = 0, 1$. Let $A_i, B_i \in \{0, 1\}$ indicate the levels of factors A and B received by unit i. We first review five factor-based regression strategies commonly used to analyse 2^2 experiments, and then clarify their respective causal interpretations.

The canonical factor-based regression takes the form $Y_i \sim 1 + A_i + B_i + A_iB_i$. Strategy (i) directly uses the coefficients of (A_i, B_i, A_iB_i), denoted by $\hat{\gamma}_0 = (\hat{\gamma}_{0,\mathrm{A}}, \hat{\gamma}_{0,\mathrm{B}}, \hat{\gamma}_{0,\mathrm{AB}})^{\mathrm{T}}$, to estimate the main effects of factors A and B, and their interaction. Strategy (ii) uses

$(\hat{\gamma}_{0,\mathrm{A}} + B_i\hat{\gamma}_{0,\mathrm{AB}}, \hat{\gamma}_{0,\mathrm{B}} + A_i\hat{\gamma}_{0,\mathrm{AB}}, \hat{\gamma}_{0,\mathrm{AB}})$ to estimate the main effects and interaction at the unit level, and then takes their respective averages to estimate the factorial effects at the population level. Define $e_{\mathrm{A}=1} = N^{-1}\sum_{i=1}^{N} A_i$ and $e_{\mathrm{B}=1} = N^{-1}\sum_{i=1}^{N} B_i$ as the empirical probabilities of factors A and B, respectively. The final estimators are $\hat{\gamma}_{\mathrm{e}} = (\hat{\gamma}_{\mathrm{e,A}}, \hat{\gamma}_{\mathrm{e,B}}, \hat{\gamma}_{\mathrm{e,AB}})^{\mathrm{T}}$, where $\hat{\gamma}_{\mathrm{e,A}} = \hat{\gamma}_{0,\mathrm{A}} + e_{\mathrm{B}=1}\hat{\gamma}_{0,\mathrm{AB}}$, $\hat{\gamma}_{\mathrm{e,B}} = \hat{\gamma}_{0,\mathrm{B}} + e_{\mathrm{A}=1}\hat{\gamma}_{0,\mathrm{AB}}$ and $\hat{\gamma}_{\mathrm{e,AB}} = \hat{\gamma}_{0,\mathrm{AB}}$. Strategy (ii) is popular in econometrics, where the estimators of the main effects, namely $\hat{\gamma}_{\mathrm{e,A}}$ and $\hat{\gamma}_{\mathrm{e,B}}$, are also known as the average partial or marginal effects (Greene, 2018). Strategy (iii) codes the factors by their signs as $A_i^{\mathrm{s}} = 2A_i - 1$ and $B_i^{\mathrm{s}} = 2B_i - 1 \in \{+1, -1\}$, and uses the coefficients from $Y_i \sim 1 + A_i^{\mathrm{s}} + B_i^{\mathrm{s}} + A_i^{\mathrm{s}}B_i^{\mathrm{s}}$, after multiplication by 2, to estimate the main effects and the interaction, respectively (Wu & Hamada, 2009; Lu, 2016). Let $\hat{\gamma}_{\mathrm{s}} = (\hat{\gamma}_{\mathrm{s,A}}, \hat{\gamma}_{\mathrm{s,B}}, \hat{\gamma}_{\mathrm{s,AB}})^{\mathrm{T}}$ denote the estimators under strategy (iii). These three strategies can simultaneously estimate the main effects and interaction via one least-squares fit.

Strategies (iv) and (v), on the other hand, focus on only the two main effects. Strategy (iv) considers two separate regressions, $Y_i \sim 1 + A_i$ and $Y_i \sim 1 + B_i$, and estimates the two main effects by the coefficients of A_i and B_i (e.g., Bertrand & Mullainathan, 2004; Eriksson & Rooth, 2014). Strategy (v) considers the additive regression $Y_i \sim 1 + A_i + B_i$ and estimates the two effects via one least-squares fit.

A factor-based regression is said to be saturated if it contains all possible interactions between the factors in addition to the constant term and main effects. The regressions under strategies (i)–(iii) are saturated, whereas those under strategies (iv) and (v) are unsaturated.

3.2. *Unifying the saturated regressions and introducing the general factorial effects*

We now unify strategies (i)–(iii) within a class of location-shifted factor-based regressions that turn out to target factorial effects under different weighting schemes. The result highlights the correspondence between the location shifts in specifying the models and the weighting schemes in defining the factorial effects.

To this end, we first formalize the notion of general factorial effects, which are central to clarifying the effective estimands under strategies (i)–(iii). Define $\tau_{\mathrm{A}|b} = \tau_{\mathrm{A}|\mathrm{B}=b} = \bar{Y}(1b) - \bar{Y}(0b)$ and $\tau_{\mathrm{B}|a} = \tau_{\mathrm{B}|\mathrm{A}=a} = \bar{Y}(a1) - \bar{Y}(a0)$ as the conditional effects of factors A and B when the level of the other factor is fixed at $b \in \{0, 1\}$ and at $a \in \{0, 1\}$, respectively. As a convention, we abbreviate the $\mathrm{A} = a$ and $\mathrm{B} = b$ in the subscripts to simply a and b when confusion is unlikely to arise. Define

$$\tau_{\mathrm{A}}(\pi_{\mathrm{B}}) = \pi_{\mathrm{B}=0}\,\tau_{\mathrm{A}|\mathrm{B}=0} + \pi_{\mathrm{B}=1}\,\tau_{\mathrm{A}|\mathrm{B}=1}, \quad \tau_{\mathrm{B}}(\pi_{\mathrm{A}}) = \pi_{\mathrm{A}=0}\,\tau_{\mathrm{B}|\mathrm{A}=0} + \pi_{\mathrm{A}=1}\,\tau_{\mathrm{B}|\mathrm{A}=1}$$

as the main effects of factors A and B under weighting schemes $\pi_{\mathrm{B}} = (\pi_{\mathrm{B}=0}, \pi_{\mathrm{B}=1})$ and $\pi_{\mathrm{A}} = (\pi_{\mathrm{A}=0}, \pi_{\mathrm{A}=1})$, respectively, with $0 \leqslant \pi_{\mathrm{A}=a}, \pi_{\mathrm{B}=b} \leqslant 1$ for $a, b = 0, 1$ and $\pi_{\mathrm{A}=0} + \pi_{\mathrm{A}=1} = \pi_{\mathrm{B}=0} + \pi_{\mathrm{B}=1} = 1$. As a convention, the subscript of the weighting scheme indicates the factor that is being marginalized out. The standard main effects correspond to $\pi_{\mathrm{A}} = \pi_{\mathrm{B}} = (1/2, 1/2)$, weighting all conditional effects equally (Dasgupta et al., 2015).

Define $\tau_{\mathrm{AB}} = \bar{Y}(11) - \bar{Y}(10) - \bar{Y}(01) + \bar{Y}(00)$ as the interaction between A and B. It satisfies $\tau_{\mathrm{AB}} = \tau_{\mathrm{A}|\mathrm{B}=1} - \tau_{\mathrm{A}|\mathrm{B}=0} = \tau_{\mathrm{B}|\mathrm{A}=1} - \tau_{\mathrm{B}|\mathrm{A}=0}$ and characterizes the difference in the conditional effects of one factor at the two levels of the other factor. Observe that $\tau_{\mathrm{A}}(\pi'_{\mathrm{B}}) - \tau_{\mathrm{A}}(\pi_{\mathrm{B}}) = (\pi'_{\mathrm{B}=1} - \pi_{\mathrm{B}=1})\tau_{\mathrm{AB}}$ and $\tau_{\mathrm{B}}(\pi'_{\mathrm{A}}) - \tau_{\mathrm{B}}(\pi_{\mathrm{A}}) = (\pi'_{\mathrm{A}=1} - \pi_{\mathrm{A}=1})\tau_{\mathrm{AB}}$, such that τ_{AB} also quantifies the difference in causal estimands between different weighting schemes. The absence of interaction, i.e., $\tau_{\mathrm{AB}} = 0$, ensures that $\tau_{\mathrm{A}}(\pi_{\mathrm{B}}) = \tau_{\mathrm{A}|\mathrm{B}=0}$ and $\tau_{\mathrm{B}}(\pi_{\mathrm{A}}) = \tau_{\mathrm{B}|\mathrm{A}=0}$ are constant across all possible weighting schemes.

Recall that $\bar{Y} = \{\bar{Y}(00), \bar{Y}(01), \bar{Y}(10), \bar{Y}(11)\}^{\mathrm{T}}$. We vectorize the main effects and the interaction as $\tau_{\pi} = \{\tau_{\mathrm{A}}(\pi_{\mathrm{B}}), \tau_{\mathrm{B}}(\pi_{\mathrm{A}}), \tau_{\mathrm{AB}}\}^{\mathrm{T}} = G_{\pi}\bar{Y}$, where $\pi = (\pi_{\mathrm{A}}, \pi_{\mathrm{B}})$ and the contrast

matrix G_π consists of row vectors $(-\pi_{B=0}, -\pi_{B=1}, \pi_{B=0}, \pi_{B=1})$, $(-\pi_{A=0}, \pi_{A=0}, -\pi_{A=1}, \pi_{A=1})$ and $(1, -1, -1, 1)$. An unbiased estimator for τ_π is $\hat{\tau}_\pi = G_\pi \hat{Y} = \{\hat{\tau}_A(\pi_B), \hat{\tau}_B(\pi_A), \hat{\tau}_{AB}\}^T$.

Let $e_{A=0} = 1 - e_{A=1}$ and $e_{B=0} = 1 - e_{B=1}$ be, respectively, the proportions of units that receive level 0 of factors A and B in the experiment. The following proposition is a numeric result and clarifies the causal interpretations of the regression estimators from strategies (i)–(iii).

PROPOSITION 1. *Under the* 2^2 *experiment, the coefficients from strategies* (i)–(iii) *satisfy:*

(i) $\hat{\gamma}_0 = \{\hat{\tau}_A(1,0), \hat{\tau}_B(1,0), \hat{\tau}_{AB}\}^T$ *with* $\pi_A = \pi_B = (1,0)$;
(ii) $\hat{\gamma}_e = \{\hat{\tau}_A(e_{B=0}, e_{B=1}), \hat{\tau}_B(e_{A=0}, e_{A=1}), \hat{\tau}_{AB}\}^T$ *with* $\pi_f = (e_{f=0}, e_{f=1})$ *for factors* $f = A, B$;
(iii) $\hat{\gamma}_s = \{\hat{\tau}_A(1/2, 1/2), \hat{\tau}_B(1/2, 1/2), \hat{\tau}_{AB}/2\}^T$ *with* $\pi_A = \pi_B = (1/2, 1/2)$.

Strategies (i)–(iii) thus yield identical estimators of τ_{AB} up to a scaling factor and yet target distinct main effects under different weighting schemes. Strategy (i) is unbiased for estimating $\tau_A(1,0) = \tau_{A|B=0}$ and $\tau_B(1,0) = \tau_{B|A=0}$ as the conditional effects when the other factor is at the baseline level. Strategy (ii) is unbiased for estimating $\tau_A(e_{B=0}, e_{B=1})$ and $\tau_B(e_{A=0}, e_{A=1})$; the average partial effects in econometrics thus weight the conditional effects by the empirical treatment probabilities. Strategy (iii) is unbiased for estimating the standard effects $\tau_A = \tau_A(1/2, 1/2)$ and $\tau_B = \tau_B(1/2, 1/2)$ that weight all conditional effects equally. This clarifies the causal interpretations of $\hat{\gamma}_0$, $\hat{\gamma}_e$ and $\hat{\gamma}_s$ from strategies (i), (ii) and (iii), respectively. In particular, $\hat{\gamma}_s$ targets the standard factorial effects regardless of whether the experiment is balanced or not.

Inspired by how transformation applied to factors allows one to obtain the moment estimators of the standard main effects directly as regression coefficients under strategy (iii), we now propose a location-shift strategy to generalize strategies (i)–(iii) and estimate τ_π with arbitrary weights $\pi = (\pi_A, \pi_B)$ via least squares. For $A_i' = A_i - \delta_A$ and $B_i' = B_i - \delta_B$ with prespecified $0 \leqslant \delta_A, \delta_B \leqslant 1$, define the location-shifted regression

$$Y_i \sim 1 + A_i' + B_i' + A_i' B_i' \tag{1}$$

with coefficients $\hat{\gamma} = (\hat{\gamma}_A, \hat{\gamma}_B, \hat{\gamma}_{AB})^T$ and robust covariance $\hat{\Psi}$ for the three nonintercept terms. Strategies (i)–(iii) are special cases: setting $(\delta_A, \delta_B) = (0,0)$ gives strategy (i); setting $(\delta_A, \delta_B) = (e_{A=1}, e_{B=1})$ is equivalent to strategy (ii) in the sense of $\hat{\gamma} = \hat{\gamma}_e$ by Proposition 2 below; and setting $(\delta_A, \delta_B) = (1/2, 1/2)$ yields strategy (iii) up to scaling factors of 2 or 4.

Recall the unbiased estimator $\hat{\tau}_\pi = G_\pi \hat{Y}$ of $\tau_\pi = G_\pi \bar{Y}$. Let $\hat{\text{cov}}(\hat{\tau}_\pi) = G_\pi \hat{V} G_\pi^T$ be the corresponding estimated covariance, where $\hat{V}$ is a conservative estimator of $\text{cov}(\hat{Y})$. The next proposition states the numeric correspondence between $\{\hat{\gamma}, \hat{\Psi}\}$ and $\{\hat{\tau}_\pi, \hat{\text{cov}}(\hat{\tau}_\pi)\}$, elucidating the design-based properties of $\hat{\gamma}$ and $\hat{\Psi}$ for general (δ_A, δ_B).

PROPOSITION 2. *Under the* 2^2 *experiment, the outputs of* (1) *satisfy* $\hat{\gamma} = \hat{\tau}_\pi$ *and* $\hat{\Psi} = \hat{\text{cov}}(\hat{\tau}_\pi) - G_\pi \,\text{diag}(N_z^{-1}) \hat{V} G_\pi^T$ *for* $\pi = (\pi_A, \pi_B)$ *with* $\pi_A = (1 - \delta_A, \delta_A)$ *and* $\pi_B = (1 - \delta_B, \delta_B)$.

Proposition 2 ensures that $\hat{\gamma}$ from (1) is unbiased for estimating τ_π with $\pi_A = (1 - \delta_A, \delta_A)$ and $\pi_B = (1 - \delta_B, \delta_B)$. Location shifts of A_i and B_i by $(\delta_A, \delta_B) = (\pi_{A=1}, \pi_{B=1})$ thus enable direct estimation of τ_π from (1) for arbitrary π. This provides the intuition behind the condition $0 \leqslant \delta_A, \delta_B \leqslant 1$ introduced earlier. Moreover, the difference between $\hat{\Psi}$ and $\hat{\text{cov}}(\hat{\tau}_\pi)$ diminishes as N goes to infinity. This enables the large-sample Wald-type inference of τ_π by using $\hat{\gamma}$ and $\hat{\Psi}$ as the point estimator and estimated covariance, respectively.

Remark 1. The classical experimental design literature focuses mostly on the standard main effects (Wu & Hamada, 2009), with equal weights on all conditional effects: $\tau_A =$

$2^{-1}(\tau_{A|B=0}+\tau_{A|B=1})$ and $\tau_B = 2^{-1}(\tau_{B|A=0}+\tau_{B|A=1})$. The standard main effects, together with balanced experiments with $N_z = N/Q$ for all $z \in \mathcal{T}$, have many advantages in practice. Corollary 1 later states a result for the 2^K experiment with a general K.

In practice, however, applications may not always value $\tau_{A|B=0}$ and $\tau_{A|B=1}$, and likewise $\tau_{B|A=0}$ and $\tau_{B|A=1}$, equally. Alternative weighting schemes based on perceived importance therefore also merit attention, and could provide more relevant summaries of the marginal effects (Finney, 1948). We give an example based on the consideration of external validity of the experimental results.

Assume that the experiment in question is a pilot study for a large-scale implementation in which one-third of the population is intended to receive level 1 of factor B marginally. Since we know that two-thirds of the population will be experiencing the effect of factor A at the baseline level of factor B, the general effect $\tau_A(2/3, 1/3) = (2/3)\tau_{A|B=0} + (1/3)\tau_{A|B=1}$ may be a better summary of the effect of factor A compared with the standard effect with equal weights. This illustrates the connection between the general weighting schemes and external validity.

When $\tau_{AB} \neq 0$, we are also interested in finding the optimal level of factor B to maximize the effect of factor A. This requires us to compare $\tau_{A|B=1}$ and $\tau_{A|B=0}$, which correspond to two special estimands $\tau_A(0, 1)$ and $\tau_A(1, 0)$.

In summary, the choice of estimand depends on the scientific question of interest. We provide the theory for the general estimand, which includes the above examples as special cases.

3.3. *Factor-based regression with unsaturated models*

Strategies (iv) and (v) concern only the main effects of factors A and B. Strategy (iv) fits two separate regressions for estimating the main effects of factors A and B. The resulting estimators equal the differences in means between $\{Y_i : f_i = 1\}$ and $\{Y_i : f_i = 0\}$ for $f = A, B$, and are biased for estimating factorial effects of the form $\tau_A(\pi_B)$ and $\tau_B(\pi_A)$ in general. We thus exclude strategy (iv) from the ensuing discussion.

Strategy (v), on the other hand, estimates the two main effects together via one additive regression. Consider a generalized version, incorporating the location-shift transformation:

$$Y_i \sim 1 + A_i' + B_i'. \tag{2}$$

We first derive the effective estimands of (2) as a pair of general factorial effects, and then state the bias-variance trade-off between (1) and (2). The result establishes the optimality of (2) for estimating arbitrary τ_π when the nuisance effect τ_{AB} indeed does not exist.

Let $\tilde{\gamma}_A$ and $\tilde{\gamma}_B$ be the coefficients of A_i' and B_i', respectively, from (2). Let $\hat{\tau}_{A|B=b}$ and $\hat{\tau}_{B|A=a}$ be the moment estimators of $\tau_{A|B=b}$ and $\tau_{B|A=a}$ for $a, b = 0, 1$, respectively.

PROPOSITION 3. *Under the 2^2 experiment, the coefficients from (2) satisfy*

$$\tilde{\gamma}_A = \tilde{\pi}_{B=0}\,\hat{\tau}_{A|B=0} + \tilde{\pi}_{B=1}\,\hat{\tau}_{A|B=1}, \quad \tilde{\gamma}_B = \tilde{\pi}_{A=0}\,\hat{\tau}_{B|A=0} + \tilde{\pi}_{A=1}\,\hat{\tau}_{B|A=1}$$

with $\tilde{\pi}_B = (\tilde{\pi}_{B=0}, \tilde{\pi}_{B=1}) = \sigma^{-1}(e_{01}^{-1} + e_{11}^{-1}, e_{00}^{-1} + e_{10}^{-1})$ *and* $\tilde{\pi}_A = (\tilde{\pi}_{A=0}, \tilde{\pi}_{A=1}) = \sigma^{-1}(e_{10}^{-1} + e_{11}^{-1}, e_{00}^{-1} + e_{01}^{-1})$, *where* $\sigma = \sum_{z\in\mathcal{T}} e_z^{-1}$.

Proposition 3 shows $\tilde{\gamma}_A$ and $\tilde{\gamma}_B$ to be the moment estimators of $\tau_A(\tilde{\pi}_B)$ and $\tau_B(\tilde{\pi}_A)$ under a specific weighting scheme that is fully determined by $(e_z)_{z\in\mathcal{T}}$ and independent of (δ_A, δ_B). Therefore, the unsaturated regression (2) no longer accommodates flexible weighting schemes even with location-shifted factors. Under balanced designs with equal treatment sizes $N_z = N/4$

for all $z \in \mathcal{T}$, $\tilde{\gamma}_{\mathrm{A}} = \hat{\tau}_{\mathrm{A}}(1/2, 1/2)$ and $\tilde{\gamma}_{\mathrm{B}} = \hat{\tau}_{\mathrm{B}}(1/2, 1/2)$ give the moment estimators of the standard main effects, and thus equal the coefficients of A_i' and B_i' from the saturated regression (1) with $\delta_{\mathrm{A}} = \delta_{\mathrm{B}} = 1/2$. This is no coincidence, but arises from the fact that the columns of the design matrix of (1) with $\delta_{\mathrm{A}} = \delta_{\mathrm{B}} = 1/2$ are mutually orthogonal, such that the deletion of $A_i'B_i'$ has no effect on the estimation of the remaining coefficients. This highlights the connection between standard effects and balanced designs from a different angle, echoing the classical principle that recommends the use of balanced designs whenever possible.

In general, $\tilde{\gamma}_{\mathrm{A}}$ and $\tilde{\gamma}_{\mathrm{B}}$ are biased for $\tau_{\mathrm{A}}(\pi_{\mathrm{B}})$ and $\tau_{\mathrm{B}}(\pi_{\mathrm{A}})$ unless $(\pi_{\mathrm{A}}, \pi_{\mathrm{B}}) = (\tilde{\pi}_{\mathrm{A}}, \tilde{\pi}_{\mathrm{B}})$ or the interaction τ_{AB} does not exist. Nevertheless, under Condition 2, they minimize the sampling variances of $\hat{\tau}_{\mathrm{A}}(\pi_{\mathrm{B}}) = \pi_{\mathrm{B}=0}\,\hat{\tau}_{\mathrm{A}|\mathrm{B}=0} + \pi_{\mathrm{B}=1}\,\hat{\tau}_{\mathrm{A}|\mathrm{B}=1}$ and $\hat{\tau}_{\mathrm{B}}(\pi_{\mathrm{A}}) = \pi_{\mathrm{A}=0}\,\hat{\tau}_{\mathrm{B}|\mathrm{A}=0} + \pi_{\mathrm{A}=1}\,\hat{\tau}_{\mathrm{B}|\mathrm{A}=1}$ over all possible π_{B} and π_{A}, respectively. In particular, the constant treatment effects ensure $\mathrm{var}(\hat{\tau}_{\mathrm{A}|\mathrm{B}=0}) = s_0(N_{00}^{-1} + N_{10}^{-1})$ and $\mathrm{var}(\hat{\tau}_{\mathrm{A}|\mathrm{B}=1}) = s_0(N_{01}^{-1} + N_{11}^{-1})$. Minimizing the variance of $\hat{\tau}_{\mathrm{A}}(\pi_{\mathrm{B}})$ is thus equivalent to having the weights proportional to the inverses of $\mathrm{var}(\hat{\tau}_{\mathrm{A}|\mathrm{B}=0})$ and $\mathrm{var}(\hat{\tau}_{\mathrm{A}|\mathrm{B}=1})$, resulting in $(\tilde{\pi}_{\mathrm{B}=0}, \tilde{\pi}_{\mathrm{B}=1})$ as defined in Proposition 3. A similar argument applies to $\tilde{\gamma}_{\mathrm{B}}$. This demonstrates the bias-variance trade-off between (1) and (2).

This concludes our discussion of the 2^2 experiment. We next extend the results to the 2^3 experiment to illustrate one additional point: with more than two factors, the factor-based regression is capable of estimating only a subset of all causally meaningful factorial effects in general, yet it regains generality in the absence of three-way interactions.

4. The 2^3 factorial experiment

4.1. *Notation and definition of the general factorial effects*

The 2^3 factorial experiment features $Q = 2^3 = 8$ treatment combinations arising from three binary factors, denoted by A, B and C. Let A_i, B_i and $C_i \in \{0, 1\}$ indicate the levels of the factors received by unit i. The eight treatment combinations consist of $\mathcal{T} = \{(abc) : a, b, c = 0, 1\}$. Let $\bar{Y}(abc)$ be the average potential outcome under treatment combination $(abc) \in \mathcal{T}$. Define the conditional effects of factors A, B and C as

$$\tau_{\mathrm{A}|bc} = \bar{Y}(1bc) - \bar{Y}(0bc), \quad \tau_{\mathrm{B}|ac} = \bar{Y}(a1c) - \bar{Y}(a0c), \quad \tau_{\mathrm{C}|ab} = \bar{Y}(ab1) - \bar{Y}(ab0),$$

respectively, with the other two factors fixed at $bc, ac, ab \in \{0, 1\}^2$. Define the conditional two-way interactions between factors A and B, factors A and C, and factors B and C as

$$\begin{aligned}
\tau_{\mathrm{AB}|c} &= \bar{Y}(11c) - \bar{Y}(10c) - \bar{Y}(01c) + \bar{Y}(00c),\\
\tau_{\mathrm{AC}|b} &= \bar{Y}(1b1) - \bar{Y}(1b0) - \bar{Y}(0b1) + \bar{Y}(0b0),\\
\tau_{\mathrm{BC}|a} &= \bar{Y}(a11) - \bar{Y}(a10) - \bar{Y}(a01) + \bar{Y}(a00),
\end{aligned}$$

respectively, with the third factor fixed at $c, b, a \in \{0, 1\}$. When there is possibility of confusion, we write out $\mathrm{A} = a$, $\mathrm{B} = b$ and $\mathrm{C} = c$ for a, b and c in the subscripts to emphasize both the factors and their respective levels; for example, $\tau_{\mathrm{A}|bc} = \tau_{\mathrm{A}|\mathrm{B}=b,\mathrm{C}=c}$ and $\tau_{\mathrm{AB}|c} = \tau_{\mathrm{AB}|\mathrm{C}=c}$. These conditional effects are the building blocks for defining the general factorial effects.

To simplify the presentation, we call a set of W numbers $(\pi_1, \ldots, \pi_W)$ a W-dimensional weighting vector if $\sum_{w=1}^{W} \pi_w = 1$ and $\pi_w \geqslant 0$; a weighting scheme is then a collection of weighting vectors with composition that will be clear from the context. Throughout this section, assume that $\pi_{\mathrm{AB}} = (\pi_{ab})_{a,b=0,1}, \pi_{\mathrm{AC}} = (\pi_{ac})_{a,c=0,1}$ and $\pi_{\mathrm{BC}} = (\pi_{bc})_{b,c=0,1}$ are some prespecified four-dimensional weighting vectors, and that $\pi_{\mathrm{A}} = (\pi_a)_{a=0,1}, \pi_{\mathrm{B}} = (\pi_b)_{b=0,1}$ and

$\pi_{\mathrm{C}} = (\pi_c)_{c=0,1}$ are some prespecified two-dimensional weighting vectors. We summarize them as $\pi = \{\pi_{\mathrm{AB}}, \pi_{\mathrm{AC}}, \pi_{\mathrm{BC}}, \pi_{\mathrm{A}}, \pi_{\mathrm{B}}, \pi_{\mathrm{C}}\} = \{\pi_{ab}, \pi_{ac}, \pi_{bc}, \pi_a, \pi_b, \pi_c : a, b, c = 0, 1\}$.

DEFINITION 1. *Under the* 2^3 *experiment, define*

$$\tau_{\mathrm{A}}(\pi_{\mathrm{BC}}) = \sum_{b,c} \pi_{bc}\, \tau_{\mathrm{A}|bc}, \quad \tau_{\mathrm{B}}(\pi_{\mathrm{AC}}) = \sum_{a,c} \pi_{ac}\, \tau_{\mathrm{B}|ac}, \quad \tau_{\mathrm{C}}(\pi_{\mathrm{AB}}) = \sum_{a,b} \pi_{ab}\, \tau_{\mathrm{C}|ab}$$

as the main effects of factors A, B *and* C *under weighting vectors* π_{BC}, π_{AC} *and* π_{AB}, *respectively; define*

$$\tau_{\mathrm{AB}}(\pi_{\mathrm{C}}) = \sum_{c=0,1} \pi_c\, \tau_{\mathrm{AB}|c}, \quad \tau_{\mathrm{AC}}(\pi_{\mathrm{B}}) = \sum_{b=0,1} \pi_b\, \tau_{\mathrm{AC}|b}, \quad \tau_{\mathrm{BC}}(\pi_{\mathrm{A}}) = \sum_{a=0,1} \pi_a\, \tau_{\mathrm{BC}|a}$$

as the two-way interactions between factors A *and* B, *factors* A *and* C, *and factors* B *and* C *under weighting vectors* π_{C}, π_{B}, *and* π_{A}, *respectively; define*

$$\tau_{\mathrm{ABC}} = \tau_{\mathrm{AB}|\mathrm{C}=1} - \tau_{\mathrm{AB}|\mathrm{C}=0} = \tau_{\mathrm{AC}|\mathrm{B}=1} - \tau_{\mathrm{AC}|\mathrm{B}=0} = \tau_{\mathrm{BC}|\mathrm{A}=1} - \tau_{\mathrm{BC}|\mathrm{A}=0} = \sum_{a,b,c} (-1)^{a+b+c+1} \bar{Y}(abc)$$

as the three-way interaction between factors A, B *and* C.

Definition 1 gives a total of $2^3 - 1 = 7$ general factorial effects, vectorized as

$$\tau_\pi = \{\tau_{\mathrm{A}}(\pi_{\mathrm{BC}}), \tau_{\mathrm{B}}(\pi_{\mathrm{AC}}), \tau_{\mathrm{C}}(\pi_{\mathrm{AB}}), \tau_{\mathrm{AB}}(\pi_{\mathrm{C}}), \tau_{\mathrm{AC}}(\pi_{\mathrm{B}}), \tau_{\mathrm{BC}}(\pi_{\mathrm{A}}), \tau_{\mathrm{ABC}}\}^{\mathrm{T}} = G_\pi \bar{Y}.$$

Following the convention from the 2^2 experiment, the subscripts of the weighting vectors indicate the factors that are being marginalized out. We refer to π as the equal weighting scheme if $\pi_{ab} = \pi_{bc} = \pi_{ac} = 1/4$ and $\pi_a = \pi_b = \pi_c = 1/2$ for all $a, b, c = 0, 1$; and we refer to π as the empirical weighting scheme if $\pi_a = N^{-1} \sum_{i=1}^N \mathcal{I}(A_i = a)$, $\pi_{ab} = N^{-1} \sum_{i=1}^N \mathcal{I}(A_i = a, B_i = b)$ and so on, equal to the empirical treatment proportions in the experiment. Although Definition 1 can be general, we focus on the following coherent weighting scheme throughout the paper.

DEFINITION 2. *A weighting scheme* π *is said to be coherent if there exists a probability distribution over* $\mathcal{T}$, *represented by* $\pi_{abc} = \mathrm{pr}(A = a, B = b, C = c)$ *for* $a, b, c = 0, 1$, *such that*

$$\begin{aligned} &\pi_a = \mathrm{pr}(A = a), && \pi_b = \mathrm{pr}(B = b), && \pi_c = \mathrm{pr}(C = c), \\ &\pi_{ab} = \mathrm{pr}(A = a, B = b), && \pi_{ac} = \mathrm{pr}(A = a, C = c), && \pi_{bc} = \mathrm{pr}(B = b, C = c). \end{aligned}$$

Coherence imposes mild restrictions on the elements in π and, building on the intuition from Remark 1, provides the causal interpretation of the general factorial effects from a thought experiment perspective. Consider a target thought experiment in which we assign unit i to combination $(abc) \in \mathcal{T}$ with probability $\mathrm{pr}\{Z_i = (abc)\} = \mathrm{pr}(A_i = a, B_i = b, C_i = c) = \pi_{abc}$. The weighting vector $\pi_{\mathrm{BC}} = (\pi_{bc})_{b,c=0,1}$ gives the marginal distribution of (B_i, C_i) and renders the weighted average $\tau_{\mathrm{A},i}(\pi_{\mathrm{BC}}) = \sum_{b,c} \pi_{bc}\, \tau_{\mathrm{A}|bc,i}$, where $\tau_{\mathrm{A}|bc,i} = Y_i(1bc) - Y_i(0bc)$, an intuitive summary of the main effect of factor A on unit i, accounting for the target treatment probabilities of factors B and C (see also Hainmueller et al., 2014; Egami & Imai, 2019; de la Cuesta et al., 2021). Averaging $\tau_{\mathrm{A},i}(\pi_{\mathrm{BC}})$ over $i = 1, \ldots, N$ yields $N^{-1} \sum_{i=1}^N \tau_{\mathrm{A},i}(\pi_{\mathrm{BC}}) = \tau_{\mathrm{A}}(\pi_{\mathrm{BC}})$ as the average effect

at the population level. The general weights as such allow for external validity beyond the actual experiment being conducted. The equal weighting scheme is coherent with $\pi_{abc} = 1/8$, implying balanced design in the thought experiment. The empirical weighting scheme is also coherent with $\pi_{abc} = e_{abc} = N_{abc}/N$.

4.2. *Factor-based regression with the saturated model*

Define $A_i' = A_i - \delta_{\mathrm{A}}$, $B_i' = B_i - \delta_{\mathrm{B}}$ and $C_i' = C_i - \delta_{\mathrm{C}}$ for prespecified $\delta = (\delta_{\mathrm{A}}, \delta_{\mathrm{B}}, \delta_{\mathrm{C}})$ with $0 \leqslant \delta_{\mathrm{A}}, \delta_{\mathrm{B}}, \delta_{\mathrm{C}} \leqslant 1$, and extend (1) to the 2^3 experiment to define

$$Y_i \sim 1 + A_i' + B_i' + C_i' + A_i'B_i' + A_i'C_i' + B_i'C_i' + A_i'B_i'C_i'. \tag{3}$$

Let $\hat{\gamma}$ and $\hat{\Psi}$ be, respectively, the coefficient vector and robust covariance of the $2^3 - 1 = 7$ nonintercept terms in (3). In this subsection we study their design-based properties, illustrating two important characteristics of factor-based regressions with more than two factors. First, saturated regressions like (3) can only recover a subset of the coherent factorial effects with weighting schemes featuring a product structure as in Definition 3 below. Second, the absence of the three-way interaction restores the generality of (3) for estimating all coherent factorial effects.

DEFINITION 3. *A coherent weighting scheme π is said to be a product weighting scheme if $\pi_{abc} = \pi_a\pi_b\pi_c$ for $a, b, c = 0, 1$.*

A product weighting scheme π is fully determined by the values of $(\pi_{\mathrm{A}=1}, \pi_{\mathrm{B}=1}, \pi_{\mathrm{C}=1})$ and implies independent factors in the corresponding thought experiment. The equal weighting scheme satisfies Definition 3 with $\pi_{\mathrm{A}=1} = \pi_{\mathrm{B}=1} = \pi_{\mathrm{C}=1} = 1/2$; the empirical weighting scheme, on the other hand, in general does not.

Let $\delta_\times$ be the product weighting scheme with $\mathrm{pr}(A_i = 1) = \delta_{\mathrm{A}}$, $\mathrm{pr}(B_i = 1) = \delta_{\mathrm{B}}$ and $\mathrm{pr}(C_i = 1) = \delta_{\mathrm{C}}$ in the corresponding thought experiment. As a convention, we use $\times$ in the subscript to indicate product weighting schemes. Let $\tau_{\delta\times} = G_{\delta\times}\bar{Y}$ be the corresponding vector of general factorial effects, $\hat{\tau}_{\delta\times} = G_{\delta\times}\hat{Y}$ its moment estimator, and $\hat{\mathrm{cov}}(\hat{\tau}_{\delta\times}) = G_{\delta\times}\hat{V}G_{\delta\times}^{\mathrm{T}}$ the estimated covariance of $\hat{\tau}_{\delta\times}$. The following proposition gives the numeric correspondence between $\{\hat{\gamma}, \hat{\Psi}\}$ and $\{\hat{\tau}_{\delta\times}, \hat{\mathrm{cov}}(\hat{\tau}_{\delta\times})\}$, elucidating the utility of (3) for inferring $\tau_{\delta\times}$.

PROPOSITION 4. *Under the 2^3 experiment, the outputs of (3) satisfy $\hat{\gamma} = \hat{\tau}_{\delta\times}$ and $\hat{\Psi} = \hat{\mathrm{cov}}(\hat{\tau}_{\delta\times}) - G_{\delta\times}\,\mathrm{diag}(N_z^{-1})\hat{V}G_{\delta\times}^{\mathrm{T}}$.*

Proposition 4 highlights the commonality and difference between the 2^2 and 2^3 experiments. On the one hand, it ensures the asymptotic equivalence between $\{\hat{\gamma}, \hat{\Psi}\}$ and $\{\hat{\tau}_{\delta\times}, \hat{\mathrm{cov}}(\hat{\tau}_{\delta\times})\}$ as N goes to infinity, and thereby allows for the large-sample Wald-type inference of $\tau_{\delta\times}$ based on (3). On the other hand, the product structure of $\delta_\times$ constrains the generality of (3), and suggests that it recovers the full vector of τ_π simultaneously if and only if $(\delta_{\mathrm{A}}, \delta_{\mathrm{B}}, \delta_{\mathrm{C}}) = (\pi_{\mathrm{A}=1}, \pi_{\mathrm{B}=1}, \pi_{\mathrm{C}=1})$ and π is a product weighting scheme. The standard effects satisfy the product structure with $\pi_{\mathrm{A}=1} = \pi_{\mathrm{B}=1} = \pi_{\mathrm{C}=1} = 1/2$ and thus admit direct estimation with $\delta_{\mathrm{A}} = \delta_{\mathrm{B}} = \delta_{\mathrm{C}} = 1/2$.

The resulting specification is equivalent to that under the $\{+1, -1\}$ coding system up to a constant scaling factor on each regressor, suggesting the specificity of the $\{+1, -1\}$ coding system to the standard effects (Wu & Hamada, 2009; Lu, 2016). The partial effects, in contrast, may or may not satisfy the product structure, and are thus not necessarily directly estimable from (3); see the Supplementary Material. This provides a useful guideline for designing and analysing factorial experiments.

One exception, however, is when the three-way interaction does not exist. The absence of τ_{ABC} renders the class of product weighting schemes equivalent to the class of coherent weighting schemes in defining the general factorial effects. We formalize the idea in Proposition 5 below. For an arbitrary weighting scheme π, let $\pi_\times$ be the product weighting scheme with $\text{pr}(A_i = 1) = \pi_{\text{A}=1}$, $\text{pr}(B_i = 1) = \pi_{\text{B}=1}$ and $\text{pr}(C_i = 1) = \pi_{\text{C}=1}$ in the corresponding thought experiment. By definition, $\pi_\times$ and π share the same marginal treatment probabilities in the underlying thought experiments, and satisfy $\pi_\times = \pi$ if π is already a product weighting scheme.

PROPOSITION 5. *Under the 2^3 experiment, if $\tau_{\text{ABC}} = 0$, then $\tau_\pi = \tau_{\pi\times}$ for all coherent π, where $\tau_\pi = G_\pi \bar{Y}$ and $\tau_{\pi\times} = G_{\pi\times}\bar{Y}$ are the vectors of general factorial effects under weighting schemes π and $\pi_\times$, respectively.*

Propositions 4 and 5 together justify the inference of τ_π from (3) with $(\delta_{\text{A}}, \delta_{\text{B}}, \delta_{\text{C}}) = (\pi_{\text{A}=1}, \pi_{\text{B}=1}, \pi_{\text{C}=1})$ for all coherent π when $\tau_{\text{ABC}} = 0$. The absence of the three-way interaction restores the generality of factor-based regressions for all coherent weighting schemes.

4.3. *Factor-based regression with an unsaturated model*

Consider an extension of (2),

$$Y_i \sim 1 + A_i' + B_i' + C_i' + A_i'B_i' + A_i'C_i' + B_i'C_i', \tag{4}$$

when only the main effects and two-way interactions are of interest, vectorized as

$$\tau_{\pi,+} = (\tau_{\text{A}}(\pi_{\text{BC}}), \tau_{\text{B}}(\pi_{\text{AC}}), \tau_{\text{C}}(\pi_{\text{AB}}), \tau_{\text{AB}}(\pi_{\text{C}}), \tau_{\text{AC}}(\pi_{\text{B}}), \tau_{\text{BC}}(\pi_{\text{A}}))^{\text{T}} = \tau_\pi \backslash \{\tau_{\text{ABC}}\}.$$

Let $\tilde{\gamma}_+$ and $\tilde{\Psi}_+$ be the coefficient vector and robust covariance of the six nonintercept terms from (4). We use $\tilde{}$ to signify outputs from unsaturated regressions, and subscript $+$ to signify quantities associated with the effects of interest throughout the paper. Let $\hat{\gamma}_+$ and $\hat{\gamma}_{\text{ABC}}$ be the coefficients of $(A_i', B_i', C_i', A_i'B_i', A_i'C_i', B_i'C_i')$ and $A_i'B_i'C_i'$ from (3), respectively, with $\hat{\gamma} = (\hat{\gamma}_+^{\text{T}}, \hat{\gamma}_{\text{ABC}})^{\text{T}}$. Proposition 6 extends Proposition 3 to the 2^3 experiment, elucidating the design-based properties of $\tilde{\gamma}_+$ via its link with $\hat{\gamma}$.

PROPOSITION 6. *Under the 2^3 experiment, we have $\tilde{\gamma}_+ = \hat{\gamma}_+ + D\hat{\gamma}_{\text{ABC}}$ with*

$$D = \left(\sum_{z\in\mathcal{T}} e_z^{-1}\right)^{-1} \left(\begin{array}{ccc|ccc} & & & \delta_{\text{B}} & \delta_{\text{C}} & 0 \\ & I_3 & & \delta_{\text{A}} & 0 & \delta_{\text{C}} \\ & & & 0 & \delta_{\text{A}} & \delta_{\text{B}} \\ \hline & & & & & \\ & 0_{3\times 3} & & & I_3 & \\ & & & & & \end{array}\right) \left(\begin{array}{c} -\sum_a e_{a00}^{-1} \\ -\sum_b e_{0b0}^{-1} \\ -\sum_c e_{00c}^{-1} \\ \hline \sum_{ab} e_{ab0}^{-1} \\ \sum_{ac} e_{a0c}^{-1} \\ \sum_{bc} e_{0bc}^{-1} \end{array}\right) - \left(\begin{array}{c} \delta_{\text{B}}\delta_{\text{C}} \\ \delta_{\text{A}}\delta_{\text{C}} \\ \delta_{\text{A}}\delta_{\text{B}} \\ \hline \delta_{\text{C}} \\ \delta_{\text{B}} \\ \delta_{\text{A}} \end{array}\right).$$

Recall that $\hat{\gamma}_+$ and $\hat{\gamma}_{\text{ABC}}$ equal the moment estimators of $\tau_{\delta\times,+}$ and τ_{ABC}, respectively, denoted by $\hat{\tau}_{\delta\times,+}$ and $\hat{\tau}_{\text{ABC}}$. The coefficients from (4) thus recover the exact moment estimator $\hat{\tau}_{\delta\times,+}$ if and only if $D = 0_6$ or $\hat{\tau}_{\text{ABC}} = 0$. The former in general entails $e_z = 1/8$ for all $z \in \mathcal{T}$ and $\delta_{\text{A}} = \delta_{\text{B}} = \delta_{\text{C}} = 1/2$, implying both balanced design and standard effects as the estimands. In particular, $e_z = 1/8$ $(z \in \mathcal{T})$ and $\delta_{\text{A}} = \delta_{\text{B}} = \delta_{\text{C}} = 1/2$ ensure that the columns of the design matrix of the saturated regression (3) are mutually orthogonal, such that deletion of any subset of the columns has no effect on the estimation of the remaining coefficients, with (4) being a

special case. This is in line with the intuition from the 2^2 case and echos the classical principle advocating the use of balanced designs whenever possible.

On the other hand, Proposition 6 implies $E(\tilde{\gamma}_+) - \tau_{\delta\times,+} = D\tau_{\text{ABC}}$, such that $\tilde{\gamma}_+$ is unbiased for $\tau_{\delta\times,+}$ as long as the nuisance effect τ_{ABC} indeed does not exist. This, together with the equivalence between τ_π and $\tau_{\pi\times}$ in the absence of τ_{ABC}, ensures the generality of (4) for estimating $\tau_{\pi,+}$. More precisely, under $\tau_{\text{ABC}} = 0$, the coefficient $\tilde{\gamma}_+$ from (4) with $(\delta_{\text{A}}, \delta_{\text{B}}, \delta_{\text{C}}) = (\pi_{\text{A}=1}, \pi_{\text{B}=1}, \pi_{\text{C}=1})$ is unbiased for $\tau_{\pi,+}$ for all coherent weighting schemes π.

Violation of the condition of no three-way interaction, on the other hand, subjects $\tilde{\gamma}_+$ to nondiminishing bias, i.e., $D\tau_{\text{ABC}}$. The intuition for the bias-variance trade-off from the 2^2 case extends here and ensures that $\tilde{\gamma}_+$ is more precise than $\hat{\gamma}_+$ under Condition 2, regardless of whether $\tau_{\text{ABC}} = 0$ or not.

5. A GENERAL THEORY FOR THE 2^K FACTORIAL EXPERIMENT

5.1. *Overview and notation*

The 2^K factorial experiment features $Q = 2^K$ treatment combinations arising from K binary factors, indexed by $k = 1, \ldots, K$. Of interest is the utility of the corresponding factor-based regressions for inferring the factorial effects of interest from the design-based perspective. We first extend the definitions of general factorial effects, the coherent weighting scheme and the product weighting scheme to the 2^K experiment, and demonstrate the utility of location-shifted regressions for recovering general effects under product weighting schemes. We then show the equivalence between the coherent and product weighting schemes under the condition of no three-way interactions. Finally, we quantify the bias-variance trade-off between the saturated and unsaturated specifications.

We use the following notation. Let $Z_{ik} \in \{0, 1\}$ denote the level of factor k received by unit i for $i = 1, \ldots, N$ and $k = 1, \ldots, K$. Let $\mathcal{F}_k = \{0, 1\} = \{0_k, 1_k\}$ be the set of possible levels of factor k, where the subscript k is used to differentiate the factors. Let $\mathcal{T} = \prod_{k=1}^K \mathcal{F}_k = \{z = (z_1, \ldots, z_K) : z_k \in \mathcal{F}_k, k = 1, \ldots, K\}$ be the set of the 2^K treatment combinations. Let $\mathcal{P}_K = \{\mathcal{K} : \emptyset \neq \mathcal{K} \subseteq [K]\}$ be the set of all nonempty subsets of $[K]$. For $\mathcal{K} \in \mathcal{P}_K$, let $z_{\mathcal{K}} = (z_k)_{k\in\mathcal{K}}$ and $z_{\overline{\mathcal{K}}} = (z_k)_{k\notin\mathcal{K}}$ index the combinations of factors in $\mathcal{K}$ and $\overline{\mathcal{K}} = [K]\backslash\mathcal{K}$, respectively, taking values from $\mathcal{F}_{\mathcal{K}} = \prod_{k\in\mathcal{K}} \mathcal{F}_k = \{0, 1\}^{|\mathcal{K}|}$ and $\mathcal{F}_{\overline{\mathcal{K}}} = \prod_{k\notin\mathcal{K}} \mathcal{F}_k = \{0, 1\}^{K-|\mathcal{K}|}$. In particular, $z_{[K]} = z \in \mathcal{T}$ and $\mathcal{F}_{[K]} = \mathcal{T}$.

5.2. *Definition of the conditional factorial effects*

Consider K types of factorial effects, quantifying the main effect of a factor when applied alone, and the two- to K-way interactions when multiple factors are applied together. We refer to them interchangeably as the first- to Kth-order factorial effects. Building on the intuition from the 2^2 and 2^3 experiments, we first define the conditional factorial effects in this subsection; then we define the general factorial effects as their respective weighted averages in the next subsection.

As a general rule, we define by induction the mth-order conditional factorial effect as the difference between two $(m-1)$th-order conditional effects for $m = 2, \ldots, K$ (Wu & Hamada, 2009). For notational simplicity, we illustrate the definition of the mth-order effects using the first m factors with $\mathcal{K} = [m]$, $z_{(m+1):K} = (z_k)_{k=m+1}^K$ and $\mathcal{F}_{(m+1):K} = \prod_{k=m+1}^K \mathcal{F}_k = \{0, 1\}^{K-m}$.

DEFINITION 4. *Let $\bar{Y}(z_1, z_{2:K})$ be the average potential outcome under $z = (z_1, z_{2:K}) \in \mathcal{T}$, and define $\tau_1(z_{2:K}) = \bar{Y}(1_1, z_{2:K}) - \bar{Y}(0_1, z_{2:K})$ as the conditional main effect of factor 1 when factors 2 to K are fixed at $z_{2:K} \in \mathcal{F}_{2:K}$.*

Given $\tau_{[m-1]}$ as the conditional $(m-1)$th-order factorial effect of factors 1 to $(m-1)$ when the rest of the factors are fixed at $z_{m:K} \in \mathcal{F}_{m:K}$ for $m = 2, \ldots, K-1$, define

$$\tau_{[m]}(z_{(m+1):K}) = \tau_{[m-1]}(1_m, z_{(m+1):K}) - \tau_{[m-1]}(0_m, z_{(m+1):K})$$

as the conditional mth-order factorial effect of factors 1 to m when the rest of the factors are fixed at $z_{(m+1):K} \in \mathcal{F}_{(m+1):K}$.

For $m = K$, define $\tau_{[K]} = \tau_{[K-1]}(1_K) - \tau_{[K-1]}(0_K)$ as the K-way interaction of factors 1 to K.

Based on Definition 4, we can obtain the explicit form of $\tau_{[m]}(z_{(m+1):K})$ in terms of the $\bar{Y}(z)$, and show that the order in which new factors are added to the combination in the induction does not matter. Definition 4 extends to general $\mathcal{K} \in \mathcal{P}_K$ by symmetry. Denote by $\tau_{\mathcal{K}}(z_{\overline{\mathcal{K}}})$ the conditional $|\mathcal{K}|$th-order factorial effect of factors in $\mathcal{K}$ when the rest of the factors are fixed at $z_{\overline{\mathcal{K}}} \in \mathcal{F}_{\overline{\mathcal{K}}}$. This gives a total of $|\mathcal{F}_{\overline{\mathcal{K}}}| = 2^{K-|\mathcal{K}|}$ conditional factorial effects for the $|\mathcal{K}|$ factors in a fixed $\mathcal{K}$. The notation from the 2^2 case is a special case where $\tau_{\mathrm{A}|b} = \tau_{\mathrm{A}}(b)$ and $\tau_{\mathrm{B}|a} = \tau_{\mathrm{B}}(a)$; likewise for $\tau_{\mathrm{A}|bc} = \tau_{\mathrm{A}}(bc)$, $\tau_{\mathrm{AB}|c} = \tau_{\mathrm{AB}}(c)$ and so on from the 2^3 case.

5.3. *Definition of the general factorial effects*

We next define the general factorial effects as weighted averages of their respective conditional counterparts. Consider $\pi(z) = \mathrm{pr}(Z_{i1} = z_1, \ldots, Z_{iK} = z_K)$ for $z = (z_1, \ldots, z_K) \in \mathcal{T}$ as the treatment probabilities under some target thought experiment. The marginal distribution of $Z_{i,\mathcal{K}} = (Z_{ik})_{k\in\mathcal{K}}$ equals $\pi_{\mathcal{K}} = \{\pi(z_{\mathcal{K}}) : z_{\mathcal{K}} \in \mathcal{F}_{\mathcal{K}}\}$ with $\pi(z_{\mathcal{K}}) = \mathrm{pr}(Z_{i,\mathcal{K}} = z_{\mathcal{K}}) = \sum_{z_{\overline{\mathcal{K}}} \in \mathcal{F}_{\overline{\mathcal{K}}}} \pi(z_{\mathcal{K}}, z_{\overline{\mathcal{K}}})$. It induces an intuitive weighting scheme for averaging over factors in $\mathcal{K}$ when defining the general factorial effect of factors in $\overline{\mathcal{K}}$. The $\pi_{\mathrm{A}} = (\pi_{\mathrm{A}=0}, \pi_{\mathrm{A}=1})$ and $\pi_{\mathrm{AB}} = (\pi_{ab})_{a,b=0,1}$ from the 2^2 and 2^3 experiments are special cases of $\pi_{\mathcal{K}}$ with $\mathcal{K} = \{\mathrm{A}\}$ and $\{\mathrm{A}, \mathrm{B}\}$, respectively. Building on the intuition from Definition 2, we call $\pi = \{\pi_{\mathcal{K}} : \mathcal{K} \in \mathcal{P}_K\}$ the coherent weighting scheme induced by the joint distribution $\{\pi(z) : z \in \mathcal{T}\}$.

DEFINITION 5. *Given a coherent weighting scheme π and conditional factorial effects $\tau_{\mathcal{K}}(z_{\overline{\mathcal{K}}})$ from Definition 4 for all $\mathcal{K} \in \mathcal{P}_K$ and $z_{\overline{\mathcal{K}}} \in \mathcal{F}_{\overline{\mathcal{K}}}$, define*

$$\tau_{\mathcal{K},\pi} = \sum_{z_{\overline{\mathcal{K}}} \in \mathcal{F}_{\overline{\mathcal{K}}}} \pi(z_{\overline{\mathcal{K}}})\, \tau_{\mathcal{K}}(z_{\overline{\mathcal{K}}})$$

as the general factorial effect of factors in $\mathcal{K}$ under π, vectorized as

$$\tau_\pi = \{\tau_{\mathcal{K},\pi} : \mathcal{K} \in \mathcal{P}_K\} = G_\pi \bar{Y}.$$

Definitions 4 and 5 together define the $2^K - 1$ general factorial effects under the coherent weighting scheme π. We refer to $\tau_{\mathcal{K},\pi}$ as the standard effect if $\pi(z_{\overline{\mathcal{K}}}) = |\mathcal{F}_{\overline{\mathcal{K}}}|^{-1} = 2^{-|\overline{\mathcal{K}}|}$ is the same for all $z_{\overline{\mathcal{K}}} \in \mathcal{F}_{\overline{\mathcal{K}}}$. We refer to $\tau_{\mathcal{K},\pi}$ as the empirical effect if $\pi(z_{\overline{\mathcal{K}}}) = N^{-1} \sum_{i=1}^N \mathcal{I}(Z_{i,\overline{\mathcal{K}}} = z_{\overline{\mathcal{K}}})$ equals the empirical proportion in the actual experiment.

5.4. *Factor-based regression with the saturated model*

Motivated by (1) for the 2^2 experiment and (3) for the 2^3 experiment, we define $Z'_{ik} = Z_{ik} - \delta_k$ and $Z'_{i,\mathcal{K}} = \prod_{k\in\mathcal{K}} Z'_{ik}$ as a location-shifted generalization for some prespecified $(\delta_k)_{k=1}^K$ with

$0 \leqslant \delta_k \leqslant 1$, and consider the saturated factor-based regression

$$Y_i \sim 1 + \sum_{k=1}^{K} Z'_{ik} + \sum_{1 \leqslant k \neq k' \leqslant K} Z'_{ik} Z'_{ik'} + \cdots + \prod_{k=1}^{K} Z'_{ik} \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z'_{i,\mathcal{K}}. \tag{5}$$

Let $\hat{\gamma}$ and $\hat{\Psi}$ be, respectively, the coefficient vector and robust covariance of the $Q-1$ nonintercept terms in (5), with elements arranged in the same order of $\mathcal{K}$ as those in τ_π. In the following we derive their utility for the Wald-type inference of τ_π.

To begin with, the notion of product weighting scheme extends naturally to the current setting as $\pi(z) = \prod_{k=1}^{K} \pi(z_k)$ and is fully determined by the values of $\{\pi(1_k)\}_{k=1}^{K}$. The equal weighting scheme for the standard effects satisfies the product structure with $\pi(1_k) = 1/2$. The empirical weighting scheme, on the other hand, may not. Building on the intuition from the 2^3 experiment, Definition 6 introduces two product weighting schemes of particular importance, arising from the estimand of interest and the location-shift parameters.

DEFINITION 6. *For an arbitrary coherent weighting scheme π, let $\pi_\times$ be the product weighting scheme with $\pi_\times(1_k) = \pi(1_k)$ for $k = 1, \ldots, K$.*

For arbitrary location-shift parameters $(\delta_k)_{k=1}^{K}$ with $0 \leqslant \delta_k \leqslant 1$, let $\delta_\times$ be the product weighting scheme with $\delta_\times(1_k) = \delta_k$ for $k = 1, \ldots, K$.

The product weighting scheme $\pi_\times$ satisfies $\pi_\times = \pi$ if π is already a product weighting scheme. The product weighting scheme $\delta_\times$ features $\delta_\times(z) = \prod_{k=1}^{K} \delta_k^{z_k}(1-\delta_k)^{1-z_k}$ for all $z \in \mathcal{T}$. Let $\tau_{\delta\times} = G_{\delta\times}\bar{Y}$ be the corresponding vector of general factorial effects, $\hat{\tau}_{\delta\times} = G_{\delta\times}\hat{Y}$ its moment estimator, and $\hat{\mathrm{cov}}(\hat{\tau}_{\delta\times}) = G_{\delta\times}\hat{V}G_{\delta\times}^{\mathrm{T}}$ the estimated covariance. The following theorem gives the numeric correspondence between $\{\hat{\gamma}, \hat{\Psi}\}$ and $\{\hat{\tau}_{\delta\times}, \hat{\mathrm{cov}}(\hat{\tau}_{\delta\times})\}$ for inferring $\tau_{\delta\times}$.

THEOREM 1. *Under the 2^K experiment, the outputs of (5) satisfy $\hat{\gamma} = \hat{\tau}_{\delta\times}$ and $\hat{\Psi} = \hat{\mathrm{cov}}(\hat{\tau}_{\delta\times}) - G_{\delta\times}\,\mathrm{diag}(N_z^{-1})\hat{V}G_{\delta\times}^{\mathrm{T}}$.*

Theorem 1 unifies the results from the 2^2 and 2^3 experiments, and justifies the utility of $\hat{\gamma}$ and $\hat{\Psi}$ from (5) for inferring τ_π when π is a product weighting scheme with $\pi(1_k) = \delta_k$ for $k = 1, \ldots, K$. Despite the constrained applicability in general, the intuition from Proposition 5 extends here and ensures the generality of (5) in the absence of three-way interactions.

Condition 3. We have $\tau_{\mathcal{K}}(z_{\overline{\mathcal{K}}}) = 0$ for all $z_{\overline{\mathcal{K}}}$ with $|\mathcal{K}| = 3$.

Condition 3 rules out the existence of three-way interactions, and hence of all m-way interactions for $3 < m \leqslant K$ by Definition 4.

THEOREM 2. *Under the 2^K experiment and Condition 3, we have $\tau_\pi = \tau_{\pi\times}$ for all coherent π, where τ_π and $\tau_{\pi\times}$ are the vectors of general factorial effects under π and $\pi_\times$, respectively.*

Theorems 1 and 2 together allow us to use (5) with $\delta_k = \pi(1_k)$ for the Wald-type inference of all τ_π with coherent π in the absence of three-way interactions. The proof of Theorem 2 further shows that the requirement of $\tau_{\mathcal{K}}(z_{\overline{\mathcal{K}}}) = 0$ for all $|\mathcal{K}| = 3$ is not only sufficient, but also necessary for $\tau_\pi = \tau_{\pi\times}$ to hold if π is coherent, but not a product weighting scheme. Thus, for Theorem 2 to hold, we cannot relax the $|\mathcal{K}| = 3$ in Condition 3 to $|\mathcal{K}| = m$ for some $m > 3$.

5.5. *Factor-based regression with unsaturated models*

Motivated by (2) for the 2^2 experiment and (4) for the 2^3 experiment, we next consider

$$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z'_{i,\mathcal{K}}, \tag{6}$$

where $\mathcal{F}_+ \subset \mathcal{P}_K$, as an unsaturated variant of (5) when only a subset of the $Q-1$ factorial effects are of interest, vectorized as $\tau_{\pi,+} = \{\tau_{\mathcal{K},\pi} : \mathcal{K} \in \mathcal{F}_+\}$. A commonly used special case is $Y_i \sim 1 + Z'_{i1} + \cdots + Z'_{iK}$, with only the first-order terms and $\mathcal{F}_+ = \{\{k\} : k = 1, \ldots, K\}$. The additive form ensures that the location-shift transformation has no effect on the estimation of the nonintercept coefficients. Another commonly used special case is $Y_i \sim 1 + \sum_{k=1}^K Z'_{ik} + \sum_{1 \leqslant k \neq k' \leqslant K} Z'_{ik} Z'_{ik'}$, with only the main effects and two-way interactions and with $\mathcal{F}_+ = \{\{k\}, \{k, k'\} : k, k' = 1, \ldots, K,\ k \neq k'\}$.

Let $\tilde{\gamma}_+$ and $\tilde{\Psi}_+$ be, respectively, the coefficient vector and robust covariance of the $|\mathcal{F}_+|$ nonintercept terms in (6). In this subsection we establish their utility for inferring $\tau_{\pi,+}$. Recall the coefficient vector $\hat{\gamma}$ of the nonintercept terms from (5); partition it into $\hat{\gamma}_+$ and $\hat{\gamma}_-$, corresponding to the coefficients of $(Z'_{i,\mathcal{K}})_{\mathcal{K} \in \mathcal{F}_+}$ and $(Z'_{i,\mathcal{K}})_{\mathcal{K} \notin \mathcal{F}_+}$, respectively. As a convention, we use $+$ and $-$ in the subscripts to signify effects included in and omitted from the unsaturated regression (6), respectively.

Let F be the $N \times Q$ design matrix of (5), concatenating columns of 1_N and $(Z'_{i,\mathcal{K}})_{i=1}^N$ for all $\mathcal{K} \in \mathcal{P}_K$. Let F_+ be the $N \times (1 + |\mathcal{F}_+|)$ design matrix of (6) and $F_- = F \backslash F_+$ the submatrix of F omitted from (6), concatenating columns of $(Z'_{i,\mathcal{K}})_{i=1}^N$ for $\mathcal{K} \notin \mathcal{F}_+$. Assume throughout that the elements in $\tilde{\gamma}_+$, $\hat{\gamma}_+$ and $\hat{\gamma}_-$ are arranged in the same relative order of $\mathcal{K}$ as those in τ_π, and likewise for the columns in F_+ and F_-.

Let $\Phi = (F_+^{\mathrm{T}} F_+)^{-1} F_+^{\mathrm{T}} F_-$ be the coefficient matrix from the columnwise regression of F_- on F_+, which is a deterministic function of $(e_z)_{z \in \mathcal{T}}$ and $(\delta_k)_{k=1}^K$; see the Supplementary Material. Let $R = F_- - F_+ \Phi$ be the corresponding residual matrix, D the submatrix of Φ without the first row, and $F_{+[,-1]}$ the submatrix of F_+ without the first column. Let P_N be the projection matrix orthogonal to 1_N, and let $Y = (Y_1, \ldots, Y_N)^{\mathrm{T}}$ be the vector of observed outcomes. The next theorem states the numeric correspondence between $\tilde{\gamma}_+$ and $\hat{\gamma}$ under the 2^K factorial experiment, generalizing Propositions 3 and 6. It is an application of Cochran's formula (Cox 2007).

THEOREM 3. *Under the 2^K experiment, the coefficients from (5) and (6) satisfy $\tilde{\gamma}_+ = \hat{\gamma}_+ + D\hat{\gamma}_-$, where $D\hat{\gamma}_- = 0$ if and only if $F_{+[,-1]}^{\mathrm{T}} P_N F_- (R^{\mathrm{T}} R)^{-1} R^{\mathrm{T}} Y = 0$. In particular, $D\hat{\gamma}_- = 0$ for all Y if $F_+^{\mathrm{T}} F_- = 0$ or $F_+^{\mathrm{T}} P_N F_- = 0$.*

Recall that $\hat{\gamma}_+$ and $\hat{\gamma}_-$ coincide with the moment estimators of $\tau_{\delta\times,+} = \{\tau_{\mathcal{K},\delta\times} : \mathcal{K} \in \mathcal{F}_+\}$ and $\tau_{\delta\times,-} = \{\tau_{\mathcal{K},\delta\times} : \mathcal{K} \notin \mathcal{F}_+\}$, respectively, denoted by $\hat{\tau}_{\delta\times,+}$ and $\hat{\tau}_{\delta\times,-}$. Theorem 3 gives two sufficient conditions for $\tilde{\gamma}_+$ to recover exactly $\hat{\tau}_{\delta\times,+}$, requiring orthogonality of F_+ and F_- either in the original form or after being centred by the column averages. These conditions do not hold in general unless the design is balanced and the factorial effects are the standard ones under the equal weighting scheme. This generalizes the intuition from the 2^2 and 2^3 cases to general K.

COROLLARY 1. *Under the 2^K experiment, $\tilde{\gamma}_+ = \hat{\gamma}_+$ if* (i) *$\delta_k = 1/2$ for all $k = 1, \ldots, K$, and* (ii) *$N_z = N/Q$ for all $z \in \mathcal{T}$.*

The balance condition (ii) in Corollary 1 can be dropped if we use the weighted least-squares fit with weights $1/N_{Z_i}$ for $i = 1, \ldots, N$. We relegate the details to the Supplementary Material and focus on the ordinary least-squares fit here.

Despite the loss of exact recovery of the moment estimator $\hat{\tau}_{\delta\times,+}$ when $D\hat{\gamma}_- \neq 0$, the intuition from the 2^2 and 2^3 experiments extends here and ensures the unbiasedness of $\tilde{\gamma}_+$ in the absence of the nuisance effects.

Condition 4. The nuisance effects are zero; that is, $\tau_{\mathcal{K},\pi} = 0$ for all $\mathcal{K} \notin \mathcal{F}_+$.

THEOREM 4. *Under the completely randomized* 2^K *experiment, the coefficients from* (6) *satisfy* $E(\tilde{\gamma}_+) = \tau_{\delta\times,+} + D\tau_{\delta\times,-}$ *and* $\mathrm{cov}(\tilde{\gamma}_+) = (I, D)G_{\delta\times}\mathrm{cov}(\hat{Y})G_{\delta\times}^{\mathrm{T}}(I, D)^{\mathrm{T}}$. *Further assume Condition* 1 *and Condition* 4 *with* $\pi = \delta_\times$. *Then* $E(\tilde{\gamma}_+) = \tau_{\delta\times,+}$, *and* $\tilde{\gamma}_+$ *is asymptotically normal with* $N\{\tilde{\Psi}_+ - \mathrm{cov}(\tilde{\gamma}_+)\} = \Delta + o_{\mathrm{p}}(1)$, *where* $\Delta = (I, D)G_{\delta\times}SG_{\delta\times}^{\mathrm{T}}(I, D)^{\mathrm{T}} \geqslant 0$.

Theorem 4 justifies the Wald-type inference of $\tau_{\delta\times,+}$ from the unsaturated specification (6) when the nuisance effects omitted indeed do not exist. The resulting $\tilde{\gamma}_+$ is both unbiased and consistent for estimating $\tau_{\delta\times,+}$, with the robust covariance $\tilde{\Psi}_+$ affording an asymptotically conservative estimator for the true sampling covariance. The proof of Theorem 1 further shows that the intercept from (6) is an unbiased estimator of a weighted average of $\bar{Y}(z)$ instead of a contrast, and is thus nonzero in general. This suggests the necessity of including the intercept in the unsaturated specification for the satisfaction of Condition 4. One limitation of (6) again lies in its requirement on the product weighting scheme. Juxtaposing Condition 3, Condition 4 and Theorem 2 ensures that the result of Theorem 4 extends to $\tau_{\pi,+}$ for all coherent π in the absence of three-way interactions.

Remark 2. The condition of constant treatment effects further ensures $\mathrm{cov}(\tilde{\gamma}_+) \leqslant \mathrm{cov}(\hat{\gamma}_+)$, such that the estimator from (6) has smaller sampling covariance than that from (5). This, together with Theorem 4, illustrates the bias-variance trade-off between the saturated and unsaturated regressions from the design-based perspective. The result, however, does not hold without the assumption of constant treatment effects; a counterexample is given in the Supplementary Material.

The assumption of no nuisance effects can never be verified exactly in practice. Extra caution is therefore needed when applying unsaturated specifications to unbalanced designs or estimands other than the standard effects. The saturated specification is, in this sense, a safer choice when the sample size permits. When the number of treatment combinations $Q = 2^K$ is large relative to the sample size N, however, the saturated regression is subject to substantial finite-sample variability, so that the unsaturated regressions are possibly more attractive for finite-sample inference. Even if the nuisance effects are not exactly zero, depending on one's belief of the data-generating process, the gain in finite-sample precision by the unsaturated regressions can still outweigh the bias as long as the omitted nuisance effects, most likely some higher-order interactions, are reasonably small, ensuring a smaller mean squared error overall.

Alternatively, lasso and ridge regression may be attractive options when the saturated regression is not possible. Indeed, the discussion so far holds with a given unsaturated specification (6). It is desirable to have a data-driven specification with both model selection and post-selection inference (Chipman et al., 1997; Espinosa et al., 2016; Egami & Imai, 2019). Although these topics have been discussed extensively under the classic linear model, analogous questions under the design-based framework remain largely unexplored. We leave such investigations to future work.

6. DISCUSSION AND RECOMMENDATIONS

We conclude this article with three practical implications of our findings in terms of the 2^K experiment. The intuition extends to the general $Q_1 \times \cdots \times Q_K$ experiment with minimal modification, as shown in the Supplementary Material. First, the definition of general factorial effects greatly broadens the range of estimands that can be considered under factorial experiments, enabling the use of flexible weighting schemes to accommodate context-specific concerns. Second, location-shifted factor-based regression affords a convenient way of recovering the moment estimators of the general factorial effects from least squares, with the corresponding robust covariance being an asymptotically conservative estimator of the true sampling covariance. This enables large-sample Wald-type inference from least-squares outputs. With more than two factors, factor-based regression is capable of estimating general factorial effects under product weighting schemes, and regains generality in the absence of three-way interactions. Third, unsaturated regressions reduce sampling variances under the condition of constant treatment effects, but are subject to nondiminishing biases when the condition of no nuisance effects is violated. Importantly, our theory is design-based without requiring any stochastic models for the potential outcomes.

We have focused on complete randomization because of its wide range of applications. Clarifying the above important issues in this basic experiment provides a proof of concept for other more complex experiments. The definitions of the general factorial effects remain unchanged, and the correspondence between the least-squares outputs and moment estimators is purely numeric, and thus holds under any randomization mechanism. The appropriateness of the Wald-type inference, however, is assignment-specific and requires modifications under different randomization mechanisms. We conjecture that the theory extends to experiments with nonconstant treatment probabilities (Mukerjee et al., 2018) if the least-squares procedure is weighted by the inverse of the treatment probability. We leave this to future research.

ACKNOWLEDGEMENT

The authors thank the reviewers and Avi Feller, Cheng Gao and Nicole Pashley for constructive comments. Zhao was supported by the Start-Up grant R-155-000-216-133 from the National University of Singapore. Ding was partially supported by the U.S. National Science Foundation.

SUPPLEMENTARY MATERIAL

Supplementary Material available at *Biometrika* online includes the results for the general $Q_1 \times \cdots \times Q_K$ factorial experiment, details omitted from the main text, and numerical examples.

REFERENCES

BERTRAND, M. & MULLAINATHAN, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *Am. Econ. Rev.* **94**, 991–1013.

BRANSON, Z., DASGUPTA, T. & RUBIN, D. B. (2016). Improving covariate balance in 2^K factorial designs via rerandomization with an application to a New York City Department of Education high school study. *Ann. Appl. Statist.* **10**, 1958–76.

CHIPMAN, H., HAMADA, M. & WU, C. F. J. (1997). A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics* **39**, 372–81.

COX, D. R. (2007). On a generalization of a result of W. G. Cochran. *Biometrika* **94**, 755–759.

Dasgupta, T., Pillai, N. & Rubin, D. B. (2015). Causal inference from 2^K factorial designs by using potential outcomes. *J. R. Statist. Soc.* B **77**, 727–53.

de la Cuesta, B., Egami, N. & Imai, K. (2021). Improving the external validity of conjoint analysis: The essential role of profile distribution. *Polit. Anal.* to appear, DOI: 10.1017/pan.2020.40.

Duflo, E., Glennerster, R. & Kremer, M. (2007). Using randomization in development economics research: A toolkit. In *Handbook of Development Economics*, T. P. Schultz & J. A. Strauss, eds., vol. 4, chap. 61. Amsterdam: Elsevier, pp. 3895–962.

Egami, N. & Imai, K. (2019). Causal interaction in factorial experiments: Application to conjoint analysis. *J. Am. Statist. Assoc.* **114**, 526–40.

Eriksson, S. & Rooth, D.-O. (2014). Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment. *Am. Econ. Rev.* **104**, 1014–39.

Espinosa, V., Dasgupta, T. & Rubin, D. B. (2016). A Bayesian perspective on the analysis of unreplicated factorial experiments using potential outcomes. *Technometrics* **58**, 62–73.

Finney, D. J. (1948). Main effects and interactions. *J. Am. Statist. Assoc.* **43**, 566–71.

Freedman, D. A. (2008). On regression adjustments to experimental data. *Adv. Appl. Math.* **40**, 180–93.

Greene, W. H. (2018). *Econometric Analysis*. Upper Saddle River, New Jersey: Pearson/Prentice Hall, 8th ed.

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

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press.

Karlan, D. & List, J. A. (2007). Does price matter in charitable giving? Evidence from a large-scale natural field experiment. *Am. Econ. Rev.* **97**, 1774–93.

Li, X. & Ding, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *J. Am. Statist. Assoc.* **112**, 1759–69.

Lu, J. (2016). On randomization-based and regression-based inferences for 2^K factorial designs. *Statist. Prob. Lett.* **112**, 72–8.

Mukerjee, R., Dasgupta, T. & Rubin, D. B. (2018). Using standard tools from finite population sampling to improve causal inference for complex experiments. *J. Am. Statist. Assoc.* **113**, 868–81.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. *Statistical Science* **5**, 465–472. translated by Dabrowska, D. M. and Speed, T. P.

Torres, C., Ogbu-Nwobodo, L., Alsan, M., Stanford, F. C., Banerjee, A., Breza, E., Chandrasekhar, A. G., Eichmeyer, S., Karnani, M., Loisel, T. et al. (2021). Effect of physician-delivered COVID-19 public health messages and messages acknowledging racial inequity on black and white adults' knowledge, beliefs, and practices related to COVID-19: A randomized clinical trial. *JAMA Network Open* **4**, e2117115.

Wu, C. F. J. & Hamada, M. (2009). *Experiments: Planning, Analysis, and Optimization*. New York: John Wiley & Sons.

Wu, J. & Ding, P. (2021). Randomization tests for weak null hypotheses in randomized experiments. *J. Am. Statist. Assoc.* to appear, DOI: 10.1080/01621459.2020.1750415.

[*Received on* 7 *January* 2021. *Editorial decision on* 27 *September* 2021]