Randomization Inference in Two-Sided Market Experiments *

Jizhou Liu
Booth School of Business
University of Chicago
jliu32@chicagobooth.edu

Azeem M. Shaikh
Department of Economics
University of Chicago
amshaikh@uchicago.edu

Panos Toulis
Booth School of Business
University of Chicago
panos.toulis@chicagobooth.edu

September 19, 2025

Abstract

Randomized experiments are increasingly employed in two-sided markets, such as buyer-seller platforms, to evaluate treatment effects from marketplace interventions. These experiments must reflect the underlying two-sided market structure in their design (e.g., sellers and buyers), making them particularly challenging to analyze. In this paper, we propose a randomization inference framework to analyze outcomes from such two-sided experiments. Our approach is finite-sample valid under sharp null hypotheses for any test statistic and maintains asymptotic validity under weak null hypotheses through studentization. Moreover, we provide heuristic guidance for choosing among multiple valid randomization tests to enhance statistical power, which we demonstrate empirically. Finally, we demonstrate the performance of our methodology through a series of simulation studies and illustrate its use in an empirical analysis of data from a network experiment.

KEYWORDS: Causal inference; Conditional randomization test; Interference; Two-sided experiments; Multiple Randomization Design.

*The second author acknowledges support from the National Science Foundation through grant SES-2419008.

1 Introduction

Randomized experiments are increasingly used to assess the impact of policies in various online domains, including social media, online marketplaces, and streaming media platforms (Kohavi et al., 2020; Athey and Luca, 2019; Thomke, 2003). These environments generate complex datasets through two-sided interactions between multiple types of actors, such as viewers and content creators or customers and suppliers. In this context, policy evaluation is complicated due to *market interference*, where the policy treatment on certain units may affect the outcomes of different units in the environment, including those never directly assigned to treatment (Blake and Coey, 2014; Toulis and Parkes, 2016). Traditional randomized experiments, designed for single-population settings, often fall short in addressing the complexities arising from interference in two-sided markets (Masoero et al., 2024; Johari et al., 2022; Wager and Xu, 2021).

To overcome these challenges, a novel class of experimental designs, termed *Multiple Randomization Designs*, has been introduced by Bajari et al. (2023) and further developed by Masoero et al. (2024), specifically for marketplace experimentation. For example, in an online marketplace with buyers and sellers, the experimenter might randomly assign half of the buyers and half of the sellers to treatment groups, applying a policy intervention (e.g., free shipping) to transactions where both the buyer and the seller are treated. Transactions where either the buyer or the seller (or both) are untreated would not receive the intervention. This design creates variation in the exposure to treatment across different transactions, which facilitates the analysis of causal effects in the presence of interference.

In this paper, we propose Fisherian-style randomization tests tailored specifically for analyzing two-sided market experiments of this kind (Fisher, 1953). We formulate two sharp null hypotheses on spillover treatment effects that are naturally aligned with the two-sided structure of these experiments. Our approach adopts the conditional randomization testing framework that restricts attention to a set of "focal units" associated with each tested hypothesis (Athey et al., 2018; Basse et al., 2019; Puelz et al., 2021). Within this framework, we propose two valid testing procedures: a permutation-based procedure, which is straightforward to implement but relies on specific symmetry conditions that may not universally hold; and an alternative procedure, valid under arbitrary designs, which requires sampling from a potentially complex conditional randomization space. For a third null hypothesis on total treatment effects, we identify a novel trade-off among multiple valid conditional randomization tests and provide heuristic guidance on designing tests with enhanced statistical power.

Next, we extend our framework to test weak null hypotheses on average spillover and total treatment effects by incorporating studentized statistics into our randomization tests. Studentization ensures the asymptotic validity of our procedures under the weak null hypothesis, while preserving finite-sample validity under the corresponding sharp null, following insights from related literature (Chung and Romano, 2013; DiCiccio and Romano, 2017a; Zhao and Ding, 2021; Wu and Ding, 2021). Our key technical contribution involves extending the proof techniques of Zhao and Ding (2021) and Masoero et al. (2024) to the context of two-sided experiments. Finally, we validate our methodology and theoretical analysis through extensive simulation studies and a synthetic data example. In particular, we empirically examine how the choice of different conditioning events in our randomization procedures impacts statistical power, and show how the empirical relationship agrees with our theory.

This paper contributes to the growing literature on randomized experiments in multi-sided online marketplaces. Bajari et al. (2023) pioneered the introduction of multiple randomization designs to account for complex spillover effects occurring within marketplaces. Masoero et al. (2024) delves deeper into such designs, adopting a Neymanian perspective that emphasizes randomization-based point estimation and proposes conservative variance estimators for statistical inference. By contrast, our work considers the Fisherian perspective that instead focuses on finite-sample exact p-values via randomization-based tests.

The rest of the paper is organized as follows. Section 2 describes the setup and notation. Section 3 presents the main results under sharp null hypotheses. Section 4 discusses the randomization tests for the weak null hypotheses. Section 5 examines the finite-sample behavior of our methods through simulations. Section 6 illustrates the proposed inference methods in an empirical application based on the experiment conducted in Comola and Prina (2021). Section 7 concludes the paper.

2 Setup and Notation

We begin by introducing the two-sided randomized design following the two-population buyer-seller example in Masoero et al. (2024). Consider a marketplace comprising I buyers and J sellers, indexed by $i = 1, \dots, I$ and $j = 1, \dots, J$, respectively. In the marketplace, buyers and sellers interact, producing an outcome of interest, such as the total number of transactions between the buyer-seller pair within a given time period. We denote this outcome between buyer i and seller j as $Y_{i,j} \in \mathbb{R}$. We denote the set of buyer-seller pairs as $\mathbb{U} = \{(i, j) : i = 1, \dots, I, j = 1, \dots, J\}$. The $I \times J$ matrix $\mathbf{Y} = [Y_{i,j}]$ captures the outcomes observed in the marketplace. To examine the impact of various interventions, researchers conduct randomized experiments at the level of individual buyer-seller pairs, using a binary treatment assignment, $W_{i,j} \in \{0, 1\}$. These interventions might include incentives such as free shipping or discounts on transaction fees. The matrix $\mathbf{W} = [W_{i,j}] \in \{0, 1\}^{I \times J}$ denotes the full treatment assignment. We use $Y_{i,j}(\mathbf{w})$ to denote the potential outcome for the pair (i, j) under treatment assignment $\mathbf{w} \in \{0, 1\}^{I \times J}$, and $\mathbf{Y}(\mathbf{w})$ for the full $I \times J$ matrix of potential outcomes under assignment $\mathbf{w}$. As usual, the observed outcomes and the potential outcomes are related to treatment assignment by the consistency relationship $\mathbf{Y} = \mathbf{Y}(\mathbf{W})$.

As highlighted by Masoero et al. (2024), treatments in marketplace experiments need not be uniformly applied across all buyers associated with a particular seller, nor across all sellers interacting with a specific buyer. In this paper, we recast the Multiple Randomization Design introduced by Bajari et al. (2023) as an *independent two-sided randomized design*, emphasizing that two independent randomizations occur separately on the two sides (buyers and sellers) of the marketplace. Following Masoero et al. (2024), a buyer-seller pair—which constitutes the unit of analysis—is considered treated only if both the buyer and the seller are individually assigned treatment. However, we deviate slightly from Definition 3.5 in Masoero et al. (2024) by relaxing the requirement that the proportion treated on each side must remain fixed. Consequently, in our setting, the full treatment matrix $\mathbf{W}$ can be represented as the outer product of the buyer treatment vector and the seller treatment vector, formalized in the following definition.

Definition 2.1 (Independent Two-Sided Randomized Design). *Let $w^B = (w_1^B, \dots, w_I^B)^\top \in \mathbb{W}^B \subseteq \{0, 1\}^I$ represent a binary treatment assignment for buyers with probability distribution p^B; and $w^S = (w_1^S, \dots, w_J^S)^\top \in \mathbb{W}^S \subseteq \{0, 1\}^J$ denote the treatment assignment for sellers with distribution p^S. An independent two-sided* *randomized design is then defined by the induced probability distribution of* $\mathbf{W} = w^B(w^S)^\top \subseteq \{0,1\}^{I\times J} = \mathbb{W}$, *as a product of the two marginal distributions, i.e.,* $p(\mathbf{W}) = p^B(w^B)p^S(w^S)$.

Example 2.1. In equation (1), we present three examples of treatment assignment matrices $\mathbf{W}_1$, $\mathbf{W}_2$, and $\mathbf{W}_3$, corresponding to buyer-side, seller-side, and two-sided designs, respectively. Treatment assignments for a single population, $\mathbf{W}_1$ and $\mathbf{W}_2$, can be equivalently represented by the assignment vectors $w^B = (0,1,0,0,1)$ and $w^S = (1,0,1,0,0)$, as defined in Definition 2.1. The assignment matrix for a two-sided randomized design, $\mathbf{W}_3$, is then defined as the element-wise product (Hadamard product) of $\mathbf{W}_1$ and $\mathbf{W}_2$, $\mathbf{W}_3 = \mathbf{W}_1 \odot \mathbf{W}_2$, or as the outer product of the two assignment vectors, $\mathbf{W}_3 = w^B(w^S)^\top$ (as specified in Definition 2.1).

$$\underset{\text{Buyer-side randomization}}{\mathbf{W}_1 = \begin{pmatrix} 0&0&0&0&0\\ 1&1&1&1&1\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 1&1&1&1&1 \end{pmatrix}} \quad \underset{\text{Seller-side randomization}}{\mathbf{W}_2 = \begin{pmatrix} 1&0&1&0&0\\ 1&0&1&0&0\\ 1&0&1&0&0\\ 1&0&1&0&0\\ 1&0&1&0&0 \end{pmatrix}} \quad \underset{\text{Two-sided randomization}}{\mathbf{W}_3 = \begin{pmatrix} 0&0&0&0&0\\ 1&0&1&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 1&0&1&0&0 \end{pmatrix}} \tag{1}$$

In Example 2.2, we demonstrate the concept of an independent two-sided experiment under complete randomization. In this design, treatment assignments for each population are uniformly distributed, ensuring a fixed fraction of treated units. Another variant of the two-sided design assigns treatments based on independent and identically distributed (i.i.d.) Bernoulli trials for each buyer and seller. As discussed later, both designs enjoy a nice property known as "design symmetry" or "exchangeability", which motivates and validates the use of permutation tests as a computationally straightforward approach.

Example 2.2 (Two-Sided Design under Complete Randomization)**.** For fixed integers, $0 < I_1 < I$ and $0 < J_1 < J$, $I_0 = I - I_1$ and $J_0 = J - J_1$, let $p^B \sim \text{Unif}(\mathbb{W}^B)$ and $p^S \sim \text{Unif}(\mathbb{W}^S)$ where $\mathbb{W}^B = \{w^B \in \{0,1\}^I : \sum_{i=1}^I w_i^B = I_1\}$ and $\mathbb{W}^S = \{w^S \in \{0,1\}^J : \sum_{i=1}^J w_i^S = J_1\}$.

This induces a uniform two-sided randomized design as follows.

$$p(\boldsymbol{W}) = \begin{cases} \begin{pmatrix} I \\ I_1 \end{pmatrix} \begin{pmatrix} J \\ J_1 \end{pmatrix} & \text{, if } \boldsymbol{W} \in \left\{w^B(w^S)^\top : w^B \in \mathbb{W}^B, w^S \in \mathbb{W}^S\right\} \\ 0 & \text{otherwise.} \end{cases}$$

Next, we introduce the concept of *local interference*, which forms the basis for the two null hypotheses of interest. Under the classical Stable Unit Treatment Value Assumption (SUTVA) with no interference (Rubin, 1972, 1980), the potential outcome can be represented by only the treatment status of a given pair, i.e. $Y_{i,j}(W_{i,j})$. However, recent literature on online marketplace experiments (Blake and Coey, 2014; Toulis and Parkes, 2016; Johari et al., 2022; Masoero et al., 2024) has pointed out that this assumption is untenable as there is likely interference due to a single seller interacting with multiple buyers or a single buyer interacting with multiple sellers. On the other hand, without any restrictions on the potential outcome function, the analysis of potential outcomes would be intractable as the treatment space is exponentially large. To make progress, we follow Masoero et al. (2024) and assume that interference may only occur between the activities of a given buyer or seller in any given buyer-seller pair.

Assumption 2.1 (Local interference)**.** Potential outcomes satisfy $Y_{i,j}(\mathbf{w}) = Y_{i,j}(\mathbf{w}')$ for all pairs (i,j) and assignments $\mathbf{w}, \mathbf{w}'$, provided the following conditions hold: (a) the assignments for the pair (i,j) coincide, such that $w_{i,j} = w'_{i,j}$; (b) the fraction of treated sellers for buyer i coincide under $\mathbf{w}$ and $\mathbf{w}'$, and (c) the fraction of treated buyers for seller j coincide under $\mathbf{w}$ and $\mathbf{w}'$.

Following Lemma 3.6 in Masoero et al. (2024) and under Assumption 2.1, we can express the potential outcomes in terms of four distinct treatment exposure cases as follows:[1]

$$Y_{i,j}(\mathbf{W}) = Y_{i,j}(w_i^B, w_j^S) = \begin{cases} Y_{i,j}(0,0) & \text{if both buyer and seller are untreated,} \\ Y_{i,j}(1,0) & \text{if buyer is treated and seller is untreated,} \\ Y_{i,j}(0,1) & \text{if buyer is untreated and seller is treated,} \\ Y_{i,j}(1,1) & \text{if both buyer and seller are treated.} \end{cases} \tag{2}$$

Certain constrasts between the above potential outcomes express spillover effects from either the buyer side or the seller side. For example, the buyer spillover effect examines the impact on a buyer-seller pair (i,j) when buyers are treated versus when no buyers are treated, holding the seller untreated. Similarly, the seller spillover effect compares the outcomes for a seller when sellers are treated versus untreated, assuming the buyer remains untreated. In Hypotheses 2.1 and 2.2, which we define below, we formalize the sharp null hypotheses to test for the existence of spillover effects from both the buyer and seller sides. Rejecting these two spillover null hypotheses provides evidence for the existence of interference, suggesting that spillover effects likely exist in the given direction. For instance, if the buyer spillover effect is non-zero, it could indicate that treated buyers' overall shopping experience is influenced by interactions with treated sellers, causing changes in behavior even when shopping from untreated sellers.

Hypothesis 2.1 (Buyer spillover)**.** The null hypothesis of no buyer spillover effects is

$$H_0^{buyer} : Y_{i,j}(0,0) = Y_{i,j}(1,0), \text{ for all } i = 1, \dots, I, j = 1, \dots, J.$$

Hypothesis 2.2 (Seller spillover)**.** The null hypothesis of no seller spillover effects is

$$H_0^{seller} : Y_{i,j}(0,0) = Y_{i,j}(0,1), \text{ for all } i = 1, \dots, I, j = 1, \dots, J.$$

Hypothesis 2.3 (Total effect)**.** The null hypothesis of no total treatment effects is given as:

$$H_0^{total} : Y_{i,j}(0,0) = Y_{i,j}(1,1), \text{ for all } i = 1, \dots, I, j = 1, \dots, J.$$

[1]To be more specific, by Assumption 2.1, the potential outcomes can be represented as $Y_{i,j}(W_{i,j}, \bar{w}_i^B, \bar{w}_j^S)$, where $\bar{w}_i^B = \frac{1}{J}\sum_{j=1}^J W_{i,j}$ and $\bar{w}_j^S = \frac{1}{I}\sum_{i=1}^I W_{i,j}$ are observed treated fractions. Since $\bar{w}_i^B = \frac{1}{J}\sum_{j=1}^J w_i^B w_j^S = w_i^B r^S$ and $\bar{w}_j^S = w_i^S r^B$, where r^S and r^B are the treated fractions of sellers and buyers, respectively, we have $Y_{i,j}(W_{i,j}, \bar{w}_i^B, \bar{w}_j^S) = Y_{i,j}(w_i^B w_j^S, w_i^B r^S, w_j^S r^B)$. Therefore, the potential outcomes are a function of only w_i^B and w_j^S whenever our testing procedures do not alter the treated fractions. This justifies the simplified notation of Equation (2).

The total treatment effect (Hypothesis 2.3) aims to capture the difference between pairs where both the buyer and the seller are treated and pairs where neither is treated. We call this a "total effect" because it encompasses both spillover and direct treatment effects. For a clearer understanding, we discuss this concept through a linear outcome model in Example 2.3 that follows.

Example 2.3. Consider a linear additive model for the potential outcomes:

$$Y_{i,j}(\mathbf{W}) = \alpha w_i^B w_j^S + \beta w_i^B + \gamma w_j^S + \epsilon_{i,j} \ .$$

Hypotheses 2.1 to 2.3 can be re-stated as: $H_0^{\text{buyer}} : \beta = 0$, $H_0^{\text{seller}} : \gamma = 0$, and $H_0^{\text{total}} : \alpha + \beta + \gamma = 0$. Here, β and γ denote the buyer and seller spillover effects, respectively, while α signifies the direct treatment effect.

Finally, we introduce Assumption 2.2, which imposes symmetry on the experimental design. This symmetry allows us to construct permutation tests that are computationally efficient, but it is not necessary to construct valid randomization procedures.

Assumption 2.2 (Design Symmetry)**.** The probability distributions of treatment assignments for buyers, p^B, and sellers, p^S, are exchangeable. Specifically, $p^B(w_1^B, \dots, w_I^B) = p^B(w_{\pi_B(1)}^B, \dots, w_{\pi_B(I)}^B)$ for all permutations $\pi_B : [I] \to [I]$ and $p^S(w_1^S, \dots, w_J^S) = p^S(w_{\pi_S(1)}^S, \dots, w_{\pi_S(J)}^S)$ for all permutations $\pi_S : [J] \to [J]$.

3 Main Method

3.1 Overview of Conditional Randomization Tests

The classical Fisher Randomization Test simulates the sampling distribution of the test statistic according to the actual treatment variation in the experiment (Imbens and Rubin, 2015, Chapter 5). While this procedure is valid for testing the sharp null hypothesis under interference, it is generally not valid for non-sharp hypotheses, which do not specify the complete schedule of potential outcomes. The null hypotheses defined in the previous section fall into this category. For instance, if we observe outcome $Y_{ij}(1, 0)$ for a treated buyer i and control seller j, then we cannot impute the outcome $Y_{ij}(0, 0)$ under H_0^{seller}—this null hypothesis is not sharp.

To address this issue, recent literature has proposed the use of conditional randomization tests that execute the resampling procedure on a subset of units, $\mathcal{U} \subseteq \mathbb{U}$ and a subset of assignments, $\mathcal{W} \subseteq \mathbb{W}$, such that the potential outcomes become fully specified (Aronow, 2012; Athey et al., 2018; Basse et al., 2019). Note also that a unit is a buyer-seller pair in our case. In the terminology of Basse et al. (2019), $\mathcal{C} = (\mathcal{U}, \mathcal{W})$ is the *conditioning event*, which may depend on the observed treatment assignment in a random way. Let $P(\mathcal{C} \mid \mathbf{W})$ denote its distribution, which is under the analyst's control. The conditional Fisher Randomization Test then randomizes treatment according to its conditional distribution:

$$P(\mathbf{W} \mid \mathcal{C}) \propto P(\mathcal{C} \mid \mathbf{W}) P(\mathbf{W}). \tag{3}$$

Such a test is finite-sample valid for a non-sharp null hypothesis, provided that the conditioning event, $\mathcal{C}$, is constructed in a way such that the potential outcomes for all units and assignments in $\mathcal{C}$ can be imputed under the null hypothesis (Basse et al., 2019, Theorem 1). Sampling from (3) may be challenging in general. However, under certain conditions on the design, $P(\mathbf{W})$, and the conditioning mechanism, $P(\mathcal{C}|\mathbf{W})$, the distribution in (3) can be reduced to a permutation distribution that is easy to sample from. We will consider this approach in the following sections.

3.2 Testing spillover effects

In this section, we develop conditional randomization tests for the spillover hypotheses H_0^{buyer} and H_0^{seller} defined above. Under symmetric designs, our randomization procedure entails straightforward permutations of treatment assignments on one side (buyer or seller), conditioned on the assignment of the other side (seller or buyer, respectively). We will begin with the buyer spillover hypothesis and describe the associated randomization procedure in detail. The analysis of the seller spillover hypothesis will be brief as it is completely symmetrical to the buyer case.

For the buyer spillover hypothesis 2.1, we propose to condition on control sellers and fix those sellers to control across randomizations to ensure that the potential outcomes can be imputed. Moreover, we will condition on the marginal number of treated buyer-sellers observed in the sample. Under Assumption 2.2, the conditional randomization distribution is exchangeable, resulting in a permutation test that can be efficiently implemented. Formally, we define the following conditioning events:

$$\mathcal{U} = \{(i,j) : w_j^{obs,S} = 0\} \text{ and } \mathcal{W} = \{v(w^{obs,S})^\top : v \in \mathbb{W}^B\} \cap \mathcal{M} \ , \tag{4}$$

where $\mathcal{M} = \{vu^\top : v \in \mathbb{W}^B, u \in \mathbb{W}^S \text{ s.t. } \sum_{i=1}^I v_i = \sum_{i=1}^I w_i^{obs,B}, \sum_{j=1}^J u_j = \sum_{j=1}^J w_j^{obs,S}\}$. Here, we use the superscript 'obs' to emphasize that we are referring to the actual treatments realized in the observed data. Thus, $\mathcal{U}$ denotes the buyer-seller pairs for which the seller part of the pair is assigned control. Set $\mathcal{W}$ denotes the assignments that these sellers in control, while $\mathcal{M}$ is the set of assignments that match the marginal number of treated buyers and sellers observed in the sample. We note that all these sets are random as they all depend on the observed treatment assignment vector, $\mathbf{W}^{obs}$. We are now ready to define the main randomization test for the buyer spillover hypothesis.

Procedure 3.1 (Test for Buyer Spillover Effect H_0^{buyer})**.** *Let $\mathcal{C} = (\mathcal{U}, \mathcal{W})$ denote the conditioning event implied by the definitions in Equation (4).*

1. *Compute the test statistic $T(\mathbf{W}^{obs} \mid \mathbf{Y}^{obs}, \mathcal{C})$ under the observed treatment assignment. For example, we can use the difference-in-means,*

$$T(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) = \frac{1}{n_1} \sum_{(i,j)\in\mathcal{U}} Y_{i,j}(1 - w_j^S)w_i^B - \frac{1}{n_0} \sum_{(i,j)\in\mathcal{U}} Y_{i,j}(1 - w_j^S)(1 - w_i^B) \ ,$$

   where

$$n_1 = \sum_{(i,j)\in\mathcal{U}} (1 - w_j^S)w_i^B, \quad n_0 = \sum_{(i,j)\in\mathcal{U}} (1 - w_j^S)(1 - w_i^B).$$

2. *For $l = 1, 2, \ldots, L$, generate $\mathbf{W}^{(l)} = w^{B,(l)}(w^{S,obs})^\top$ where $w^{B,(l)}$ is a random permutation of $w^{B,obs}$, and compute $T(\mathbf{W}^{(l)} \mid \mathbf{Y}^{obs}, \mathcal{C})$.*

3. *Reject H_0^{buyer}, denoted by $\phi(\mathbf{W}^{obs}; \mathcal{C})$, if*

$$\frac{1}{L+1}\left[1+\sum_{l=1}^{L}\mathbb{I}\left\{T(\mathbf{W}^{(l)} \mid \mathbf{Y}^{obs}, \mathcal{C}) > T(\mathbf{W}^{obs} \mid \mathbf{Y}^{obs}, \mathcal{C})\right\}\right] \leq \alpha.$$

Figure 1 is a graphical depiction of Procedure 3.1 for testing the buyer spillover effect. The shaded parts of the array in the figure illustrates that the test conditions on the set of sellers that were not treated. Procedure 3.1 then randomizes the treatments of buyers (i.e., rows in the array). Under certain assumptions, such randomization is equivalent to permutations as explained in the following remark.

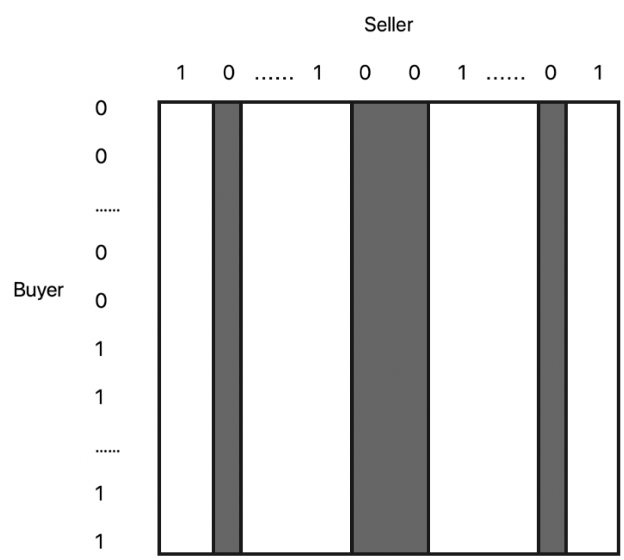

Figure 1: Graphical illustration of the conditioning event for Procedure 3.1. The shaded area illustrates that the procedure conditions on control sellers (columns), and then randomizes the treatments on buyers (rows).

Remark 3.1. Under Assumption 2.2, we can show that the permutation of the buyer treatments in step 2 leads to a valid permutation test by extending the equivariance result of Basse et al. (2024). In particular, following the definitions in Equation (4), we can show that a permutation in the assignment of buyer-seller treatment does not affect the selected buyer-seller pairs in $\mathcal{U}$. Moreover, such permutation also implies an analogous permutation of the treatment exposures of these pairs, a property known as equivariance. We give the full details of the proof in the Appendix. ■

Remark 3.2. If Assumption 2.2 does not hold, the permutation of the buyer treatments in step 2 no longer leads to a valid randomization test. In such settings, we need to modify step 2 of Procedure 3.1 as follows:

2'. For $l = 1, 2, \ldots, L$, draw $\mathbf{W}^{(l)} \sim p(\mathbf{W} \mid \mathcal{C})$ and compute $T(\mathbf{W}^{(l)} \mid \mathbf{Y}^{obs}, \mathcal{C})$. Here, $\mathbf{W}^{(l)}$ is defined as $w^{B,(l)}(w^{S,obs})^\top$, where

$$w^{B,(l)} \sim p^B\left(w \mid \sum_{i=1}^{I} w_i = \sum_{i=1}^{I} w_i^{B,obs}\right).$$

See Appendix A.1 for more details and the proof. That is, we repeatedly sample from the buyer-side design, conditioning on the treated fraction in the observed buyer-side assignment. However, sampling from such conditional distribution—e.g., through rejection sampling—may be computationally challenging depending on the particular experimental design. ■

Remark 3.3 (Seller spillover hypothesis)**.** Testing for the seller spillover hypothesis is completely analogous to Procedure 3.1. The idea is to condition on control buyers, and change the definitions in Equation (4) into

$$\mathcal{U} = \{(i, j) : w_j^{obs,B} = 0\} \text{ and } \mathcal{W} = \{w^{obs,B} u^\top : u \in \mathbb{W}^S\} \cap \mathcal{M}.$$

Operationally, we may simply transpose the buyer-seller array and execute Procedure 3.1. ■

The validity of our proposed randomization tests in Procedure 3.1 follows by adapting Theorem 1 of Basse et al. (2019) in the setting of two-sided experiments. We provide the statement of the validity result in the following theorem.

Theorem 3.1. *Consider an independent two-sided randomized design (Definition 2.1) where Assumption 2.1 holds. Let $\mathcal{C}$ be a buyer spillover conditioning event, as defined in Equation (4).*

1. *Under Assumption 2.2, the testing procedure $\phi(\mathbf{W}^{obs}; \mathcal{C})$, as defined in Procedure 3.1, is valid. Specifically, for any significance level $\alpha \in (0, 1)$, the expectation $E[\phi(\mathbf{W}^{obs}; \mathcal{C}) \mid \mathcal{C}] \leq \alpha$ holds under the null hypothesis H_0^{buyer} (or H_0^{seller}).*
2. *If Step (b) of Procedure 3.1 is replaced with the one in Remark 3.2, the validity of $\phi(\mathbf{W}^{obs}; \mathcal{C})$ is maintained without assuming Assumption 2.2.*

3.3 Testing Total effects

3.3.1 Testing Procedure with k-Block Conditioning Event

In this section, we proceed to discuss the testing procedure for total effects. We propose a randomization procedure that relies on a *k-Block Conditioning Event*. The idea is to split focal units into k-by-k blocks, with the distinctive arrangement ensuring that units within a specific block do not share rows or columns with units from other blocks. Furthermore, every unit within a block shares the same treatment status for both buyers and sellers.

Figure 2 illustrates an example of a k-block conditioning event corresponding to the shaded diagonal blocks. The treatment assignments within each block are either $(w_i^B, w_j^S) = (0, 0)$ or $(w_i^B, w_j^S) = (1, 1)$, with no overlap of blocks across columns or rows. Such arrangement is important as it allows the use of efficient permutation procedures under symmetric two-sided randomized designs. In other words, under symmetric treatment assignment designs, such as complete randomization and Bernoulli trials, our proposed tests involve permuting the treatment assignments of focal units across blocks.

Towards testing the total effect, let $\mathcal{I}^{(k)}(\mathbf{W}^{obs})$ denote a partition of the buyer set $[I]$ into k equal-sized non-overlapping subsets with all buyers having the same treatment status within each subset.[2] That is,

$$\mathcal{I}^{(k)}(\mathbf{W}^{obs}) = \{\mathcal{I}_s : 1 \leq s \leq I/k\}, \tag{5}$$

such that for all $1 \leq s, s' \leq I/k$:

(i) $|\mathcal{I}_s| = k$, and $\cup_s \mathcal{I}_s = [I]$, and $\mathcal{I}_s \cap \mathcal{I}_{s'} = \emptyset$.

(ii) $w_i^{obs,B} = w_{i'}^{obs,B}$ for all $i, i' \in \mathcal{I}_s$.

Similarly, we define $\mathcal{J}^{(k)}(\mathbf{W}^{obs})$ as a partition of the seller set $[J]$ into k-sized subsets with sellers having the same treatment status within each subset. Next, define

$$\mathcal{U}^{(k)}(\mathbf{W}^{obs}) = \big\{\{(i,j) : w_i^{obs,B} = w_j^{obs,S}, i \in \mathcal{I}_s, j \in \mathcal{J}_s\}, 1 \leq s \leq \min(I/k, J/k)\big\}, \tag{6}$$

where $\mathcal{I}_s, \mathcal{J}_s$ are defined above. That is, $\mathcal{U}^{(k)}(\mathbf{W}^{obs})$ is the collection of buyer-seller pairs constructed from the cross-product of $\mathcal{I}^{(k)}(\mathbf{W}^{obs})$ and $\mathcal{J}^{(k)}(\mathbf{W}^{obs})$, making sure that all buyers and sellers have the same treatment status in the group. Figure 2 below illustrates this construction. In the figure, the set $\mathcal{U}^{(k)}(\mathbf{W}^{obs})$ therefore corresponds to the diagonal blocks in the shaded area of the buyer-seller array. Note how these blocks are non-overlapping, they partition the sets $[I]$ and $[J]$, and, within each block, the individual treatments of every buyer and seller are identical.

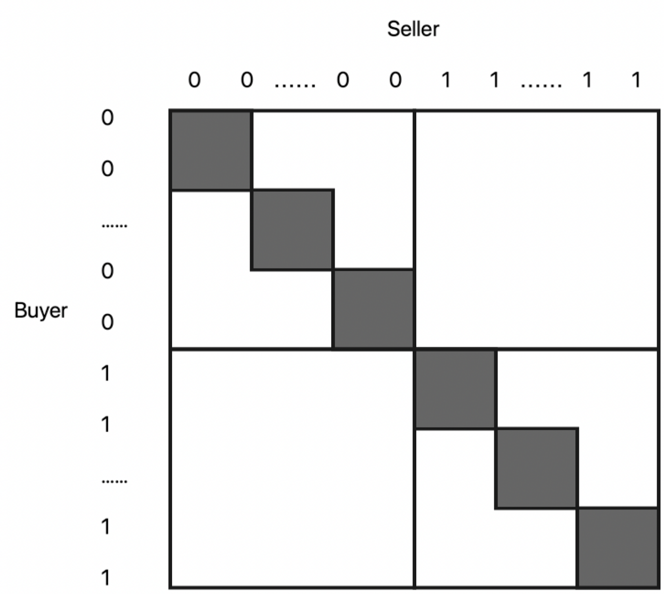

Figure 2: Graphical illustration of the conditioning event for Procedure 3.2. The shaded area illustrates that the procedure permutes the treatment assignments across diagonal blocks.

[2]We assume I and $\sum_{i=1}^{I} w_i^{obs,B}$ are both divisible by k for simplicity. When this is not the case, our methods remain valid, and the framework can be extended accordingly, though at the cost of more cumbersome notation.

Definition 3.1 (k-Block Conditioning Event)**.** *Given treatment assignment $\mathbf{W}^{obs}$, a k-block conditioning event is defined as $\mathcal{C}^{(k)} = (\mathcal{U}^{(k)}(\mathbf{W}^{obs}), \mathcal{W}^{(k)}(\mathbf{W}^{obs}))$, where $\mathcal{U}^{(k)}(\mathbf{W}^{obs})$ is defined in Equation (6) and $\mathcal{W}^{(k)}$ is a subset of treatment assignments defined as follows:*

$$\mathcal{W}^{(k)}(\mathbf{W}^{obs}) = \left\{\mathbf{W} : w_i^B = w_j^S \text{ for all } (i,j) \in \mathcal{U}^{(k)}(\mathbf{W}^{obs})\right\} \cap \mathcal{M} ,$$

where $\mathcal{M}$ is defined in Equation (4) as the set of assignments that preserve the treated fraction in each population.

We are now ready to define our main randomization procedure for testing the total null hypothesis.

Procedure 3.2 (Test Total Effect H_0^{total})**.** *Let $\mathcal{C}^{(k)} = (\mathcal{U}^{(k)}(\mathbf{W}^{obs}), \mathcal{W}^{(k)}(\mathbf{W}^{obs}))$ denote the k-conditioning event from Definition 3.1.*

1. *Compute test statistics $T(\mathbf{W}^{obs} \mid \mathbf{Y}^{obs}, \mathcal{C}^{(k)})$ under the observed treatment assignment. For example, we can use the difference-in-means:*

$$T(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}^{(k)}) = \frac{1}{n_{1,k}} \sum_{(i,j)\in\mathcal{U}^{(k)}(\mathbf{W}^{obs})} Y_{i,j} w_j^S w_i^B - \frac{1}{n_{0,k}} \sum_{(i,j)\in\mathcal{U}^{(k)}(\mathbf{W}^{obs})} Y_{i,j}(1-w_j^S)(1-w_i^B) ,$$

   where

$$n_{1,k} = \sum_{(i,j)\in\mathcal{U}^{(k)}(\mathbf{W}^{obs})} w_j^S w_i^B, \quad n_{0,k} = \sum_{(i,j)\in\mathcal{U}^{(k)}(\mathbf{W}^{obs})} (1-w_j^S)(1-w_i^B).$$

2. *For $l = 1, 2, \ldots, L$, obtain $\mathbf{W}^{(l)} = w^{B,(l)}(w^{S,(l)})^\top$ where $w^{B,(l)}, w^{S,(l)}$ are block-wise random permutation of $w^{B,obs}, w^{S,obs}$ given by*

$$w^{B,(l)} = \left\{ w^B \in \mathbb{W}^B \mid \begin{cases} (w_i^B)_{i\in\mathcal{I}_s} = (w_i^{B,obs})_{i\in\mathcal{I}_{\pi(s)}}, & \text{if } s \leq \min(I/k, J/k), \\ (w_i^B)_{i\in\mathcal{I}_s} = (w_i^{B,obs})_{i\in\mathcal{I}_s}, & \text{otherwise.} \end{cases} \right\}$$

$$w^{S,(l)} = \left\{ w^S \in \mathbb{W}^S \mid \begin{cases} (w_j^S)_{j\in\mathcal{J}_s} = (w_j^{B,obs})_{j\in\mathcal{J}_{\pi(s)}}, & \text{if } s \leq \min(I/k, J/k), \\ (w_j^B)_{i\in\mathcal{J}_s} = (w_j^{B,obs})_{j\in\mathcal{J}_s}, & \text{otherwise.} \end{cases} \right\}$$

   where π is a permutation on $\{1, \ldots, \min(I/k, J/k)\}$. Compute the randomized statistic $T(\mathbf{W}^{(l)} \mid \mathbf{Y}^{obs}, \mathcal{C}^{(k)})$.

3. *Reject H_0^{total}, denoted by $\phi(\mathbf{W}^{obs}; \mathcal{C}^{(k)}) = 1$, if*

$$\frac{1}{L+1}\left[1 + \sum_{l=1}^{L} \mathbb{I}\left\{T(\mathbf{W}^{(l)} \mid \mathbf{Y}^{obs}, \mathcal{C}^{(k)}) \geq T(\mathbf{W}^{obs} \mid \mathbf{Y}^{obs}, \mathcal{C}^{(k)})\right\}\right] \leq \alpha.$$

Intuitively, to test H_0^{total}, Procedure 3.2 conditions on a k-block conditioning event —e.g., the k diagonal blocks shown in Figure 2— and permutes the block treatments $(0,0)$, $(1,1)$ at the block level. We note that the diagonal structure in the blocks is not essential and alternative constructions can be valid as long as the blocks are non-overlapping in neither the buyer nor the seller dimension.

The following result shows that Procedure 3.2 is valid under randomized designs that satisfy local interference and design symmetry.

Theorem 3.2. *Consider an independent two-side ranomdized design (Definition 2.1) where Assumption 2.1 holds. Let $\mathcal{C}^{(k)}$ be a k-block conditioning event as in Definition 3.1. Suppose also that Assumption 2.2 holds. Then, $\phi(\mathbf{W}^{obs}; \mathcal{C}^{(k)})$ defined in Procedure 3.2 is finite-sample valid, such that $\mathbb{E}[\phi(\mathbf{W}^{obs}; \mathcal{C}^{(k)}) \mid \mathcal{C}^{(k)}] \leq \alpha$ for any finite I, J.*

In the case of non-symmetric designs where Assumption 2.2 does not hold, we can still construct valid randomization tests, albeit in a more complex form. Specifically, suppose without loss of generality that $I = J$ and the k-partitioning on two populations are identical, i.e. $\mathcal{I}^{(k)}(\mathbf{W}^{obs}) = \mathcal{J}^{(k)}(\mathbf{W}^{obs})$. Then, the permutation in step 2 of Procedure 3.2 with sampling from the conditional

$$w^{B,(l)} \sim p^B \left(w \mid \sum_{i=1}^{I} w_i^B = \sum_{i=1}^{I} w_i^{B,obs}, w_i^B = w_j^B \text{ for all } i, j \in \mathcal{I}_s \text{ for all } 1 \leq s \leq I/k \right). \tag{7}$$

and setting $\mathbf{W}^{(l)} = w^{B,(l)}(w^{B,(l)})^\top$. Despite its simplicity, this approach may be complex to implement since sampling from the conditional (7) could be computationally challenging under arbitrary non-symmetric designs.

3.3.2 Selecting k Based on Power Considerations

Theorem 3.2 establishes that Procedure 3.2 is valid for any block size k. However, the statistical power of the procedure likely depends on a complex trade-off relating to the value of k. A smaller value of k leads to using more blocks and thus is beneficial for the test's power since it increases the support of the randomization distribution. However, it also leads to smaller block sizes, and thus it may decrease power due to the reduced sample size. Conversely, a higher value of k leads to the reverse effect.

To illustrate further, consider a completely randomized two-sided design following Example 2.2, with I_1 and J_1 treated rows and columns, respectively. For the sake of simplicity, let's assume the observed data matrix is square with half of the rows and columns assigned to the treatment group, such that $I = J = 2n$ and $I_1 = I_0 = J_1 = J_0 = n$, with n/k being an integer. Under these conditions, the total number of unique treatment assignments —i.e., the support of the randomization distribution— in a k-block conditioning event equals $C(2n/k, n/k) = O(2^{2n/k}/\sqrt{2n/k})$, where $C(\cdot, \cdot)$ represents the binomial coefficient. Moreover, the sample size of focal units is $|\mathcal{U}^{(k)}(\mathbf{W}^{obs})| = 2nk$. Thus, as k increases, the sample size increases but the support of the randomization distribution decreases, which have opposing effects on the test power. Conversely, as k decreases, the total sample size decreases but the support of the randomization distribution increases.

To navigate this trade-off, we can leverage the results of Puelz et al. (2021), who established a lower bound for the power of general conditional randomization tests under some plausible assumptions, which we detail in Appendix A.2. Let $\phi_{n,k}$ denote the power of a conditional randomization test in our setting. The following proposition provides the bound under complete randomization, as in Example 2.2:

Proposition 3.1. *Under the assumptions of Theorem 3 (Appendix A.2) in Puelz et al. (2021) and the complete randomization design of Example 2.2, the power of Procedure 3.2 satisfies:*

$$\phi_{|\mathcal{U}|,|\mathcal{W}|} \geq \frac{1}{1 + Ae^{-a\tau\sqrt{2nk}}} - O\left(C(2n/k, n/k)^{-0.5+\delta}\right) - \epsilon,$$

for any small δ and sufficiently large $|\mathcal{W}|$, where $A, a > 0$ are constants.

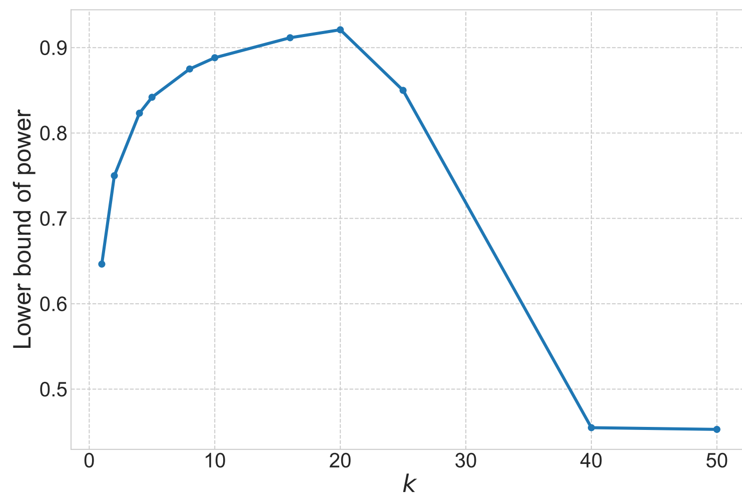

Figure 3: Graphical illustration of the lower bound on power for varying k with $n = 400$, $A = 1$, and $a = \tau = 0.01$.

In Figure 3, we illustrate the lower bound on power as a function of the block size k, highlighting the trade-off between sample size and the number of unique treatment assignments as k increases. While the plot identifies the optimal k that maximizes power, in practice, the parameters of the lower bound function are unknown. Consequently, the optimal k cannot be directly determined using Proposition 3.1. However, a heuristic approach can be used to select the largest feasible k based on a predetermined maximum power threshold. Based on Proposition 3.1, $O\left(C(2n/k, n/k)^{-0.5+\delta}\right)$ controls the maximum power of the test, while $2nk$ determines how rapidly the power function reaches its maximum (sensitivity). Therefore, to optimize the detection of treatment effects, we recommend selecting the largest feasible block size k based on a predetermined maximum power. For example, setting this power to 0.95 ensures the maximal utilization of observations to detect the minimal size of the treatment effect. In practice, we can choose k to satisfy $C(2n/k, n/k) \approx \frac{1}{(1-\beta)^2}$, where β is the predetermined power level. For example, consider the completely randomized two-sided design discussed before, where choosing a block size $k = n/6$ results in $|\mathcal{W}| = C(12, 6) = 924$, achieving a maximum power approximately equal to 0.967.

4 Randomization Tests for Weak Null Hypotheses

In this section, we focus on testing weak null hypotheses on the treatment effects considered so far. We specifically focus on a weak null for the buyer spillover hypothesis, as the other two hypotheses of interest follow similar principles. For testing total effects, refer to the discussion in Appendix B.2.3.

Hypothesis 4.1 (Weak buyer spillover)**.** The weak null hypothesis of no buyer spillover effects is:

$$H_0^{wb} : \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{i,j}(0,0) = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{i,j}(1,0)\ . \tag{8}$$

There has been a growing number of works on the study of randomization tests under weak null hypotheses similar to the definition in Equation (8); see, for example, (Chung and Romano, 2013; Li and Ding, 2017; Canay et al., 2017; DiCiccio and Romano, 2017a,b; Zhao and Ding, 2021; Wu and Ding, 2021). This literature has pointed out that randomization tests based on sharp nulls are not necessarily valid under the weak null. However, properly studentized test statistics can restore the validity of sharp randomization tests even under the weak null. Following this approach, we propose a permutation test employing a studentized test statistic for testing H_0^{wb} assuming a completely randomized design.

Procedure 4.1 (Test Weak Null H_0^{wb})**.** *Let $\mathcal{C} = (\mathcal{U}, \mathcal{W})$ denote the conditioning event implied by the definitions in Equation (4).*

1. *Compute the studentized test statistics $T^{WB}(\mathbf{W}^{obs} \mid \mathbf{Y}^{obs}, \mathcal{C})$ under the observed treatment assignment,*

$$T^{WB}(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) = \frac{T^{B}(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})}{\sqrt{V^{B}(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})}}\ ,$$

where

$$T^{B}(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) = \frac{1}{n_1}\sum_{(i,j)\in\mathcal{U}} Y_{i,j}(1-w_j^S)w_i^B - \frac{1}{n_0}\sum_{(i,j)\in\mathcal{U}} Y_{i,j}(1-w_j^S)(1-w_i^B)\ ,$$

and

$$V^{B}(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) = \frac{1}{I_1(I_1-1)}\sum_{i=1}^{I} w_i^B\left(\bar{Y}_i^B - \hat{Y}(1)\right)^2 + \frac{1}{I_0(I_0-1)}\sum_{i=1}^{I}(1-w_i^B)\left(\bar{Y}_i^B - \hat{Y}(0)\right)^2,$$

with sample mean $\hat{Y}(z) = \sum_{i:w_i^B=z}\bar{Y}_i^B/I_z$ and $I_z = \sum_{i=1}^{I}\mathbf{I}\{w_i^B = z\}$ for $z \in \{0,1\}$, and $\bar{Y}_i^B = \frac{1}{J_0}\sum_{j=1}^{J} Y_{i,j}(1-w_j^S)$.

2. *For $l = 1, 2, \ldots, L$, generate $\mathbf{W}^{(l)} = w^{B,(l)}(w^{S,obs})^\top$ where $w^{B,(l)}$ is a random permutation of $w^{B,obs}$, and compute $T^{WB}(\mathbf{W}^{(l)} \mid \mathbf{Y}^{obs}, \mathcal{C})$.*

3. *Reject H_0^{wb} if*

$$\frac{1}{L+1}\left[1 + \sum_{l=1}^{L}\mathbb{I}\left\{T^{WB}(\mathbf{W}^{(l)} \mid \mathbf{Y}(\mathbf{W}^{(l)}) \geq T^{WB}(\mathbf{W}^{obs} \mid \mathbf{Y}^{obs}, \mathcal{C})\right\}\right] \leq \alpha.$$

The following theorem provides the conditions for the asymptotic validity of Procedure 4.1 under the weak null.

Theorem 4.1. *Consider an indepdent two-side randomized design under complete randomization where Assumption 2.1 holds. Additionally, assume the following:*

1. *Balance: $I/I_0, I/I_1, J/J_0, J/J_1 \leq C_1$.*

2. Boundedness: for all buyer-seller pairs (i,j) and all $b,s \in \{0,1\}$, $|Y_{i,j}(b,s)| \leq C_2$.

For any level $\alpha \in (0,1)$, the testing procedure in Procedure 4.1 satisfies:

$$\lim_{I,J\to\infty} P\left\{E\left[\mathbb{I}\left\{T^{WB}(\mathbf{W} \mid \mathbf{Y}^{obs},\mathcal{C}) \geq T^{WB}(\mathbf{W}^{obs} \mid \mathbf{Y}^{obs},\mathcal{C})\right\}\right] \leq \alpha\right\} \leq \alpha \ ,$$

under the null hypothesis H_0^{wb}, where the expectation is with respect to $p(\mathbf{W} \mid \mathcal{C})$.

The proof of Theorem 4.1 utilizes a common strategy in the literature, demonstrating that under the studentized test statistic the sampling distribution is stochastically dominated by the randomization distribution. The analysis begins by considering the case conditional on a fixed seller-side randomization w^S, then reducing the problem to a single-population scenario. In this setup, unit-level potential outcomes for a given buyer are defined as the average over untreated sellers, with randomization solely applied to the buyer side. Leveraging existing results (Zhao and Ding, 2021; Wu and Ding, 2021), the studentized test statistic yields a randomization distribution that converges asymptotically to a standard normal distribution that stochastically dominates the sampling distribution. A primary challenge lies in extending this result to the unconditional scenario as $I, J \to \infty$. A crucial step is to observe that the conditional sampling distribution is centered at the point

$$\tau(w^S) = \frac{1}{IJ_0}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{i,j}(1,0)(1-w_j^S) - \frac{1}{IJ_0}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{i,j}(0,0)(1-w_j^S),$$

whereas the weak null hypothesis concerns τ, given by

$$\tau = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{i,j}(0,0) - \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{i,j}(1,0).$$

Following the methodological approach of Masoero et al. (2024), we demonstrate the concentration of the conditional mean and variance, thereby substantiating the validity.

Remark 4.1. The procedure for assessing total effects should adhere to that described in Procedure 4.1, substituting $\bar{Y}_i^B$ with the block average. Section 3.3.2 advises selecting the maximal permissible block size k, constrained by a predetermined maximum power level. However, for the weak null hypothesis, a large number of blocks is essential to ensure the accuracy of the variance estimator $V^B(\mathbf{W} \mid \mathbf{Y},\mathcal{C})$. The simulation results from Table 4 illustrate deteriorating size control with increasing block size, necessitating a smaller block size k for testing the weak null hypothesis. ■

5 Simulations

In this section, we examine the finite-sample behavior of our randomization tests for both total and spillover effects. Particularly, we compare the performance of randomization test against Neymanian style inference proposed in Masoero et al. (2024). Suppose that there are $3n$ units in each population, i.e. $I = J = 3n$, and the design is a completely randomized two-sided design with n units assigned to treatment group in both populations. Following Section 7.1 of Masoero et al. (2024), potential outcomes are generated as follows.

$$Y_{ij}(w,h) \overset{\text{ind}}{\sim} \begin{cases} F_0(\cdot) & \text{if } w=h=0, \\ F_0(\cdot)+F_B(\cdot) & \text{if } w=1, h=0, \\ F_0(\cdot)+F_S(\cdot) & \text{if } w=0, h=1, \\ F_0(\cdot)+F_1(\cdot) & \text{if } w=h=1, \end{cases}$$

where F_ℓ are normal distributions such that $F_\ell(\cdot)=\mathcal{N}\left(\mu_\ell,\sigma_\ell^2\right)$ for $\ell\in\{0,B,S,1\}$. In such a setting, we consider the problem of testing row spillover and total treatment effects, and compare our testing procedures from Procedure 3.1, 3.2 and 4.1 with Masoero et al. (2024). Under the sharp null hypothesis, parameters are set as follows: $\mu_0=\mu_S=\mu_B=\mu_1=0$, $\sigma_0=0.2$ and $\sigma_B=\sigma_S=\sigma_1=0$. Under the weak null hypothesis, μ_l are set to be the same as under sharp null, but variances are set to be different: $\sigma_0=0.2, \sigma_B=\sigma_1=0.4$ and $\sigma_S=0$. Under the alternative hypotheses, we set the parameters in a similar way but set the spillover effect to be 0.01 and total effects to be 0.02, i.e. $\mu_B=0.01, \mu_1=0.02$.

		Under H_0			Under H_1		
Procedure	n	FRT	Neymanian	FRT adjusted	FRT	Neymanian	FRT adjusted
Test Buyer Spillover Effects (Procedure 3.1)	10	5.02%	0.68%	4.78%	8.88%	3.84%	8.84%
	20	4.82%	3.04%	5.26%	19.96%	11.92%	20.46%
	30	5.00%	3.66%	4.90%	42.42%	28.24%	41.94%
	40	4.78%	2.86%	5.02%	58.50%	52.40%	57.76%
	50	4.72%	2.42%	4.62%	81.42%	73.66%	81.26%
	100	4.74%	3.28%	4.84%	100.00%	100.00%	100.00%
Test Total Effects (Procedure 3.2)	10	4.78%	1.82%	5.02%	5.64%	3.66%	5.50%
	20	4.84%	1.24%	4.94%	10.16%	23.32%	10.26%
	30	4.84%	0.92%	4.70%	18.92%	57.52%	17.44%
	40	4.96%	1.60%	5.12%	30.58%	86.56%	30.04%
	50	5.06%	1.56%	4.82%	42.58%	97.90%	40.34%
	100	5.24%	1.28%	4.92%	95.38%	100.00%	93.82%

Table 1: Rejection probabilities under sharp null and alternative hypothesis

Table 1 displays the rejection probabilities under the sharp null and alternative hypotheses, computed from 5,000 Monte Carlo replications with the p-values approximated by 500 independent permutations of the treatment vector in each replication. "FRT" stands for the testing procedure based on sharp null hypotheses as in Procedure 3.1 and 3.2. "FRT adjusted" stands for the studentized randomization tests as in Procedure 4.1. "Neymanian" stands for t-tests using conservative estimators of variances from Masoero et al. (2024). The block size for Procedure 3.2 is set to be $k=\lfloor n/4\rfloor$, resulting in $|\mathcal{W}|=C(12,4)=495$ and a maximum power approximately equal to 0.955. The results show that the rejection probabilities of both "FRT" and "FRT adjusted" are universally around 0.05 under the null hypothesis, which verifies the finite-sample exactness of our tests across all the designs. Meanwhile, the Neymanian inference method is conservative as expected by Masoero et al. (2024). Under the alternative hypotheses, the rejection probabilities of our methods are higher than Neymanian method when testing spillover effects, but lower when testing total effects. The power gain by Procedure 3.1 likely comes from the test's exactness, whereas the power loss from Procedure 3.2 is due to the loss in sample size used for calculating test statistics.

		Block Size k							
Hypothesis	Method	1	2	4	5	10	20	25	50
H_0^{total}	FRT	5.22%	5.40%	5.34%	5.04%	4.90%	4.98%	4.88%	0.84%
	FRT adjusted	5.24%	4.84%	4.88%	4.86%	5.26%	4.98%	5.02%	1.08%
H_1^{total}	FRT	13.20%	19.56%	35.90%	43.46%	69.40%	91.06%	95.10%	6.98%
	FRT adjusted	13.32%	20.78%	35.92%	41.80%	69.28%	89.68%	93.56%	13.62%

Table 2: Rejection probabilities under different block sizes with $n = 100$

We further investigate the robustness of the randomization inference by examining the influence of block size on both the size and power of the testing procedures. Table 2 presents our analysis of rejection probabilities under the sharp null and alternative hypotheses, following the same Monte Carlo setup in Table 1. We observe that the statistical power of the methods under investigation increases monotonically with the enlargement of the block size. Specifically, when block size is set at $k = 25$, where the randomization space encompasses $|\mathcal{W}| = C(12, 4) = 495$, near-maximum power is achieved. However, a further increase in block size to $k = 50$ results in a diminished randomization space of $|\mathcal{W}| = C(6, 2) = 15$, leading to a notable decline in test power. These results are consistent with the pattern observed in the graphical illustration of the theoretical lower bound in Figure 3.

5.1 Weak null hypotheses

Table 3 presents the rejection probabilities under the weak null and alternative hypotheses, also computed from 5,000 Monte Carlo replications and 500 independent permutations of the treatment vector in each replication. In Table 3, the block size is fixed to be 2, so that the number of block is large enough to guarantee the precision of variance estimation.

In Table 4, we characterize the effect of block size on the performance of total effect tests by following the same simulation study in Table 2. Our findings here illuminate a distinctive pattern: under the weak null hypothesis, the standard Fisher Randomization Test (FRT) tends to over-reject, whereas the adjusted FRT remains valid.

Notably, the spillover effect tests are conservative, while the total effect tests are close to exactness when the block size is small and invalid when the block size is too large. The asymptotic exactness of total effect tests is potentially due to the effect of "sampling-based versus design-based uncertainty" (Abadie et al. (2020)). The data generating process can be thought of as subsampling the focal units $|\mathcal{U}|$ from the finite population of the entire matrix. In the spillover effect tests, $|\mathcal{U}| = O(n^2)$, while the total effect tests have $|\mathcal{U}| = O(n)$. Therefore, the spillover effect tests are more subject to the design-based finite population framework, while the total effect tests are closer to the sampling-based super population framework.

Procedure	n	Under H_0 FRT	Neymanian	FRT adjusted	Under H_1 FRT	Neymanian	FRT adjusted
Test Buyer Spillover Effects (Procedure 4.1)	10	9.12%	0.6%	4.24%	14.08%	3.34%	8.20%
	20	9.60%	2.24%	4.30%	16.20%	1.94%	8.56%
	30	9.54%	1.26%	3.98%	22.84%	8.46%	12.86%
	40	8.50%	1.80%	3.36%	35.90%	15.00%	21.06%
	50	9.62%	2.06%	4.20%	46.74%	23.14%	31.24%
	100	8.64%	1.26%	3.00%	93.86%	82.98%	87.70%
Test Total Effects (Procedure 4.1)	10	13.84%	0.82%	8.22%	15.24%	1.36%	8.16%
	20	12.14%	1.18%	6.04%	16.48%	5.36%	10.64%
	30	11.90%	1.30%	5.70%	19.78%	10.48%	13.36%
	40	11.46%	0.84%	5.16%	24.76%	22.16%	16.74%
	50	11.56%	1.18%	5.52%	30.74%	38.04%	21.30%
	100	11.44%	0.86%	4.84%	67.32%	97.82%	53.56%

Table 3: Rejection probabilities under weak null and alternative hypothesis

Hypothesis	Method	Block Size k: 1	2	4	5	10	20	25	50
H_0^{total}	FRT	11.22%	12.24%	11.68%	12.10%	13.42%	12.76%	13.24%	1.52%
	FRT adjusted	5.00%	5.32%	5.04%	5.82%	5.74%	8.12%	8.44%	2.04%
H_1^{total}	FRT	14.94%	18.48%	24.58%	27.40%	40.62%	60.34%	66.48%	6.78%
	FRT adjusted	7.32%	9.60%	14.36%	16.90%	27.44%	44.46%	54.44%	9.34%

Table 4: Rejection probabilities under different block sizes with $n = 100$ under weak null

6 Data Application

In this section, we illustrate our methodology using a data application based on the empirical study of Comola and Prina (2021). The data come from a randomized field experiment that provided access to formal savings accounts to a random sample of 915 households in 19 villages near Pokhara, Nepal. The outcome of interest is a measure of intervention-induced changes in financial risk-sharing networks, constructed from adjacency matrices $\mathbf{A}^t$, where $t \in \{0, 1\}$ denotes the baseline and endline survey periods. Specifically, a semi row-standardized version of the adjacency matrix is computed for each period, denoted by $\mathbf{G}^t = \{G^t_{i,j}\}_{1\le i,j\le 915}$. The outcome is then defined as the change in standardized network links: $Y_{i,j} = G^1_{i,j} - G^0_{i,j}$.

While this setting may not at first appear to involve a two-sided market, we construct a two-sided structure using household-level covariates.[3] Specifically, we partition the households into two groups: those that experienced a death or livestock shock, and those that did not. Households experiencing shocks are considered to be in greater financial need and are thus interpreted as "buyers" of informal loans, whereas those without shocks are treated as potential "sellers" or lenders. Based on this structure, we test three

[3]The original dataset includes only household-pair-level covariates. However, we are able to recover household-level covariates by examining all pairwise records involving a given household.

hypotheses using our randomization procedures: buyer spillover, seller spillover, and total effects.

First, the buyer spillover hypothesis is defined as $H_0^{\text{buyer}} : Y_{i,j}(0,0) = Y_{i,j}(1,0)$, testing whether the treatment status of the buyer (shocked household) affects the link, holding the seller untreated. Second, the seller spillover hypothesis is defined as $H_0^{\text{seller}} : Y_{i,j}(0,0) = Y_{i,j}(0,1)$, testing whether the treatment of the seller (non-shocked household) affects the link when the buyer is untreated. Finally, the total effect hypothesis, $H_0^{\text{total}} : Y_{i,j}(0,0) = Y_{i,j}(1,1)$, tests whether there is any effect when both the buyer and seller are treated.

To test H_0^{buyer} and H_0^{seller}, we apply the one-sided permutation procedure described in Procedure 3.1. For H_0^{total}, we leverage the fact that the network is censored—survey data record only within-village links. This allows us to use the block-wise permutation procedure in Procedure 3.2, where each village defines a permutation block.

Table 5 reports p-values from our randomization-based tests alongside those from a standard two-sample t-test. We include the t-test for comparison, noting that the more principled Neyman-style inference methods (Masoero et al., 2024) are not directly applicable due to the censored nature of the data. In contrast, our approach remains valid by conditioning on the observed household pairs. Across all three hypotheses, we fail to reject the nulls, suggesting that providing savings accounts to shocked households did not significantly alter their risk-sharing relationships.

Our findings complement those of Comola and Prina (2021), who reported significant treatment effects on network behavior in the full population. Our results highlight that such effects do not appear to be driven by the subset of households experiencing shocks. This aligns with prior evidence that such effects can be highly heterogeneous: some find limited usage of randomly provided bank accounts among poor people (Dupas et al., 2016), while others report high take-up but no clear impact on aggregate expenditure, assets, or income (Prina, 2015). Together, these findings suggest that the effectiveness of savings accounts depends critically on how households interact with them—mere access may not be sufficient to alter financial behaviors.

Hypothesis	FRT	FRT adjusted	Two-sample t-test
H_0^{buyer}	74.90%	74.46%	48.74%
H_0^{seller}	80.52%	79.52%	40.93%
H_0^{total}	65.68%	65.92%	76.57%

Table 5: p-values based on various testing methods

7 Conclusion

Motivated by recent advances in experimentation within online marketplaces, this paper develops randomization-based inference procedures for two-sided market experiments. Our proposed tests are finite-sample valid under sharp null hypotheses of no treatment effect, and asymptotically valid for weak null hypotheses on average treatment effects. Additionally, we offer practical guidance for test implementation based on power considerations. Promising directions for future research include extending our framework to accommodate more complex interference structures beyond the cross-product (buyer-seller) interactions examined here. Another extension would incorporate explicit market-clearing mechanisms, such as pricing rules.

A Futher Details

A.1 Details for Remark 3.2

Consider the conditional density function $p(\mathbf{W} \mid \mathcal{C})$ with $\mathbf{W} = w^B(w^S)^\top$. With slight abuse of notation, we have

$$
\begin{aligned}
p(\mathbf{W} \mid \mathcal{C}) = \frac{p(\mathbf{W}, \mathcal{C})}{p(\mathcal{C})} &= \frac{p\left(w^B, w^S = w^{S,obs}, \sum_{i=1}^I w_i^B = \sum_{i=1}^I w_i^{B,obs}\right)}{p\left(w^S = w^{S,obs}, \sum_{i=1}^I w_i = \sum_{i=1}^I w_i^{B,obs}\right)} \\
&= \frac{p\left(w^B, \sum_{i=1}^I w_i^B = \sum_{i=1}^I w_i^{B,obs}\right) p^S(w^{S,obs})}{p^S(w^{S,obs}) p\left(\sum_{i=1}^I w_i = \sum_{i=1}^I w_i^{B,obs}\right)} \\
&= p^B\left(w^B \mid \sum_{i=1}^I w_i^B = \sum_{i=1}^I w_i^{B,obs}\right),
\end{aligned}
$$

where the final term represents the conditional distribution of w^B on the event $\{\sum_{i=1}^I w_i^B = \sum_{i=1}^I w_i^{B,obs}\}$.

A.2 Details for Proposition 3.1

Assumption A.1. Let $n = |\mathcal{U}|$ and $m = |\mathcal{Z}|$. Let the randomization distribution and the null distribution be denoted, respectively, by

$$
T(\mathbf{W} \mid \mathbf{Y}(\mathbf{W}), \mathcal{C}) \sim \hat{F}_{1,n,m}, \text{ and } T(\mathbf{W} \mid \mathbf{Y}^{obs}, \mathcal{C}) \sim \hat{F}_{0,n,m}, \text{ where } \mathbf{W} \sim P(\mathbf{W} \mid \mathcal{C}). \tag{9}
$$

Suppose that for any fixed $n > 0$:

(A.1) There exist continuous cdfs $F_{1,n}$ and $F_{0,n}$ such that $\hat{F}_{1,n,m}$ and $\hat{F}_{0,n,m}$ in (25) are the empirical distribution functions over m independent samples from $F_{1,n}$ and $F_{0,n}$, respectively.

(A.2) There exists $\sigma_n > 0$, and a continuous cdf F, such that $F_{0,n}(t) = F(t/\sigma_n)$, for all $t \in \mathbb{R}$.

(A.3) The treatment effect (e.g., spillover contrast) is additive, that is, there exists a fixed $\tau \in \mathbb{R}$ such that $F_{1,n}(t) = F_{0,n}(t - \tau)$, for all $t \in \mathbb{R}$.

B Proof of Main Results

B.1 Proof of Theorem 3.1 and 3.2

The proof is structured into two parts. The first part establishes the validity of the randomization test that samples from conditional randomization space by demonstrating that the test statistics are imputable under the conditioning events. The second part confirms the validity of the permutation test by proving the randomization hypotheses.

B.1.1 Proof for Randomization Test

The proof is a direct application of Theorem 1 of Basse et al. (2019) by verifying that any test statistic restricted on $\mathcal{C}$ is imputable, i.e. $T(\mathbf{W}' \mid \mathbf{Y}(\mathbf{W}'), \mathcal{C}) = T(\mathbf{W}' \mid \mathbf{Y}(\mathbf{W}), \mathcal{C})$. For clarity in exposition, we first show imputability for the difference-in-mean estimators in Procedure 3.1 and 3.2, and then repeat the same argument for a general test statistics. Note that, in Procedure 3.1

$$T(\mathbf{W} \mid \mathbf{Y}(\mathbf{W}), \mathcal{C}) = \frac{1}{n_1} \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{i,j}(w_i^B, w_j^S)(1 - w_j^S) w_i^B - \frac{1}{n_0} \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{i,j}(w_i^B, w_j^S)(1 - w_j^S)(1 - w_i^B) \ .$$

Consider $\mathbf{W}' = w^{B,\prime}(w^{S,\prime})^\top \in \mathcal{W}$. Under the conditioning set of assignments $\mathcal{W} = \{v(w^S)^\top : v \in \mathbb{W}^B\} \cap \mathcal{M}$, we have $w^{S,\prime} = w^S$. Then,

$$\begin{aligned}
&T(\mathbf{W}' \mid \mathbf{Y}(\mathbf{W}), \mathcal{C}) \\
&= \frac{1}{n_1} \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{i,j}(w_i^B, w_j^S)(1 - w_j^S) w_i^{B,\prime} - \frac{1}{n_0} \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{i,j}(w_i^B, w_j^S)(1 - w_j^S)(1 - w_i^{B,\prime}) \\
&= \frac{1}{n_1} \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{i,j}(w_i^B, 0)(1 - w_j^S) w_i^{B,\prime} - \frac{1}{n_0} \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{i,j}(w_i^B, 0)(1 - w_j^S)(1 - w_i^{B,\prime}) \ .
\end{aligned}$$

Under H_0^{buyer}, we have

$$Y_{i,j}(1,0) = Y_{i,j}(0,0) \text{ for all } i \in [I], j \in [J] \ .$$

Therefore,

$$\begin{aligned}
T(\mathbf{W}' \mid \mathbf{Y}(\mathbf{W}), \mathcal{C}) &= \frac{1}{n_1} \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{i,j}(w_i^{B,\prime}, 0)(1 - w_j^S) w^{B,\prime} - \frac{1}{n_0} \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{i,j}(w^{B,\prime}, 0)(1 - w_j^S)(1 - w_i^{B,\prime}) \\
&= T(\mathbf{W}' \mid \mathbf{Y}(\mathbf{W}'), \mathcal{C}) \ .
\end{aligned}$$

Next, we show imputability for Procedure 3.2. Note that

$$T(\mathbf{W} \mid \mathbf{Y}(\mathbf{W}), \mathcal{C}^{(k)}) = \frac{1}{n_{1,k}} \sum_{(i,j) \in \mathcal{U}} Y_{i,j}(w_i^B, w_j^S) w_j^S w_i^B - \frac{1}{n_{0,k}} \sum_{(i,j) \in \mathcal{U}} Y_{i,j}(w_i^B, w_j^S)(1 - w_j^S)(1 - w_i^B) \ .$$

Under the conditioning event, we have $w_i^B = w_j^S$ for all $(i,j) \in \mathcal{U}$. Therefore, we have

$$T(\mathbf{W} \mid \mathbf{Y}(\mathbf{W}), \mathcal{C}^{(k)}) = \frac{1}{n_{1,k}} \sum_{(i,j) \in \mathcal{U}} Y_{i,j}(w_i^B, w_i^B) w_i^B - \frac{1}{n_{0,k}} \sum_{(i,j) \in \mathcal{U}} Y_{i,j}(w_i^B, w_i^B)(1 - w_i^B) \ .$$

Then, by $H_0^{total} : Y_{i,j}(1,1) = Y_{i,j}(0,0)$ for all $i \in [I], j \in [J]$,

$$\begin{aligned}
T(\mathbf{W}' \mid \mathbf{Y}(\mathbf{W}), \mathcal{C}^{(k)}) &= \frac{1}{n_{1,k}} \sum_{(i,j) \in \mathcal{U}} Y_{i,j}(w_i^B, w_i^B) w_i^{B,\prime} - \frac{1}{n_{0,k}} \sum_{(i,j) \in \mathcal{U}} Y_{i,j}(w_i^B, w_i^B)(1 - w_i^{B,\prime}) \\
&= \frac{1}{n_{1,k}} \sum_{(i,j) \in \mathcal{U}} Y_{i,j}(w_i^{B,\prime}, w_i^{B,\prime}) w_i^{B,\prime} - \frac{1}{n_{0,k}} \sum_{(i,j) \in \mathcal{U}} Y_{i,j}(w_i^{B,\prime}, w_i^{B,\prime})(1 - w_i^{B,\prime})
\end{aligned}$$

$$= T(\mathbf{W}' \mid \mathbf{Y}(\mathbf{W}'), \mathcal{C}^{(k)}) \ .$$

Now, we show imputability for general test statstics in Procedure 3.1. Note that

$$\begin{aligned}
T(\mathbf{W}' \mid \mathbf{Y}(\mathbf{W}), \mathcal{C}) &= T(\mathbf{W}' \mid \{\mathbf{Y}_{i,j}(w_i^B, w_j^S)\}_{(i,j):w_j^S=0}, \{w^{S,\prime} = w^S\}, \mathcal{C}) \\
&= T(\mathbf{W}' \mid \{\mathbf{Y}_{i,j}(w_i^B, 0)\}_{(i,j):w_j^S=0}, \{w^{S,\prime} = 0\}, \mathcal{C}) \\
&= T(\mathbf{W}' \mid \{\mathbf{Y}_{i,j}(w_i^{B,\prime}, 0)\}_{(i,j):w_j^{S,\prime}=0}, \{w^{S,\prime} = 0\}, \mathcal{C}) \qquad (\text{by } H_0^{total}) \\
&= T(\mathbf{W}' \mid \mathbf{Y}(\mathbf{W}'), \mathcal{C}) \ .
\end{aligned}$$

In Procedure 3.2,

$$\begin{aligned}
T(\mathbf{W}' \mid \mathbf{Y}(\mathbf{W}), \mathcal{C}) &= T_k(\mathbf{W}' \mid \{\mathbf{Y}_{i,j}(w_i^B, w_j^S)\}_{(i,j)\in\mathcal{U}}, \{w_i^{B,\prime} = w_j^{S,\prime}\}, \mathcal{C}^{(k)}) \\
&= T_k(\mathbf{W}' \mid \{\mathbf{Y}_{i,j}(w_i^B, w_i^B)\}_{(i,j)\in\mathcal{U}}, \{w_i^{B,\prime} = w_j^{S,\prime}\}, \mathcal{C}^{(k)}) \qquad (w_i^B = w_j^S \text{ for } (i,j)\in\mathcal{U}) \\
&= T_k(\mathbf{W}' \mid \{\mathbf{Y}_{i,j}(w_i^{B,\prime}, w_i^{B,\prime})\}_{(i,j)\in\mathcal{U}}, \{w_i^{B,\prime} = w_j^{S,\prime}\}, \mathcal{C}^{(k)}) \qquad (\text{by } H_0^{total}) \\
&= T^B(\mathbf{W}' \mid \mathbf{Y}(\mathbf{W}'), \mathcal{C}^{(k)}) \ .
\end{aligned}$$

B.1.2 Proof for Permutation Test

Apparently, $\mathbf{W}^{(l)}$ belongs to the conditional event, i.e $\mathbf{W}^{(l)} \in \mathcal{W}$ for all $l = 1, \dots, L$ under both Procedure 3.1 and 3.2, meaning that the test statistics are imputable under the null hypothesis and the test procedures are well defined. Then, the validity of those procedures rests upon the *randomization hypothesis* $\mathbf{W}^{(l)} \mid \mathcal{C} \overset{d}{=} \mathbf{W}^{obs} \mid \mathcal{C}$ for all $l = 1, \dots, L$. To see this, by definition, $\sum_{l=1}^{L} E[\phi(\mathbf{W}^{(l)}; \mathcal{C}) \mid \mathcal{C}] \le \alpha L$. By randomization hypothesis,

$$\sum_{l=1}^{L} E[\phi(\mathbf{W}^{(l)}; \mathcal{C}) \mid \mathcal{C}] = \sum_{l=1}^{L} E[\phi(\mathbf{W}^{obs}; \mathcal{C}) \mid \mathcal{C}] = L \cdot E[\phi(\mathbf{W}^{obs}; \mathcal{C}) \mid \mathcal{C}] \le \alpha L \ .$$

Therefore, $E[\phi(\mathbf{W}^{obs}; \mathcal{C}) \mid \mathcal{C}] \le \alpha$.

Then, it suffices to show that randomization hypothesis holds under these two procedures. Let π^B denote a permutation operator on the buyer side assignments, and π^{block} denote the block-wise permutation. For Procedure 3.1 (b), by exchangeability of p^B, we have $(\pi_B w^B, w^S) \overset{d}{=} (w^B, w^S)$. Thus, we have $(\pi_B w^B, w^S) \mid w^S \overset{d}{=} (w^B, w^S) \mid w^S$, which implies $\pi_B \mathbf{W} \mid \mathcal{C} \overset{d}{=} \mathbf{W} \mid \mathcal{C}$. Similarly, for Procedure 3.2 (b), by exchangeability of p^B and p^S, we have $(\pi^{block} w^B, \pi^{block} w^S) \overset{d}{=} (w^B, w^S)$. Without loss of generality, we assume $I = J$ and $\mathcal{I}^{(k)}(\mathbf{W}^{obs}) = \mathcal{J}^{(k)}(\mathbf{W}^{obs}) = \{(1, \dots, k), (k+1, \dots, 2k), \dots\}$, i.e. the focal units consist of k by k diagonal blocks from a square matrix. Thus, we have

$$\begin{aligned}
&(\pi^{block} w^B, \pi^{block} w^S) \mid \{\pi^{block} w^B = \pi^{block} w^S\} \\
&= (\pi^{block} w^B, \pi^{block} w^S) \mid \{w^B = w^S\} \\
&\overset{d}{=} (w^B, w^S) \mid \{w^B = w^S\} \ ,
\end{aligned}$$

which implies $\pi^{block} \mathbf{W} \mid \mathcal{C} \overset{d}{=} \mathbf{W} \mid \mathcal{C}$. Note that for both Procedure 3.1 (b) and 3.2 (b), the conditioning event $\mathcal{C}$ has a $\mathcal{W}$ that is more restrictive conditioning set than $\{w^S\}$ (or $\{w^B = w^S\}$). For example, for Procedure

3.1 (b), $\mathcal{W} = \{v(w^S)^\top : v \in \mathbb{W}^B\} \cap \mathcal{M}$ imposes a restriction of $\mathcal{M}$ (fixed treatment fraction) in addition to $\{w^S\}$. For Procedure 3.2 (b), in addition to $\mathcal{M}$, $\mathcal{W}$ requires that $w_i^S = w_j^S$ for all $i, j \in \mathcal{I}_s$ for all s (same for w^B), i.e. treatment status should be the same within each partition. That said, those restrictions do not invalidate our argument above, because it is easy to show that $\pi w \in \mathcal{M}$ if and only if $w \in \mathcal{M}$ (same for the additional requirement by Procedure 3.2 (b)). Therefore, they are omitted for readability.

■

B.2 Proof of Section 4: Randomization Tests for Weak Null Hypotheses

Following Zhao and Ding (2021), the validity of Procedure 4.1 can be established if we can show the sampling distribution of $T^{WB}(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})$ is stochastically dominated by its randomization distribution for almost all sequences of W as $I, J \to \infty$. To begin with, we provide the following definition of the potential outcomes averaged over the untreated sellers:

$$\bar{Y}_i^B(z) = \frac{1}{J_0}\sum_{j=1}^{J} Y_{i,j}(z,0)(1-w_j^S) \text{ for } z \in \{0,1\}.$$

B.2.1 Sampling Distribution

First, we work on the sampling distribution. Let τ denote the average buyer spillover effect:

$$\tau = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{i,j}(1,0) - \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{i,j}(0,0) \; .$$

Let $v_I(w^S)$ denote the finite-population variance of $T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})$ under seller-side assignment w^S:

$$v_I(w^S) = \frac{1}{I_1}S_1^2(w^S) + \frac{1}{I_0}S_0^2(w^S) - \frac{1}{I}S_\tau^2(w^S) \; ,$$

where $S_z^2(w^S) = (I-1)^{-1}\sum_{i=1}^{I}(\bar{Y}_i^B(z) - \bar{Y}_\cdot^B(z))^2$ and $S_\tau^2(w^S) = (I-1)^{-1}\sum_{i=1}^{I}(\bar{Y}_i^B(1) - \bar{Y}_i^B(0) - \tau)^2$. Note that, for any fixed w^S,

$$\begin{aligned}
&\sup_{t\in\mathbb{R}}\left| P\left\{\frac{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) - \tau}{\sqrt{V^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})}}\sqrt{\frac{v_I(w^S) + S_\tau^2(w^S)/I}{v_I(w^S)}} \le t \mid w^S\right\} - \Phi(t)\right| \\
&= \sup_{t\in\mathbb{R}}\left| P\left\{\frac{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) - \tau}{\sqrt{v_I(w^S)}}\frac{\sqrt{v_I(w^S) + S_\tau^2(w^S)/I}}{\sqrt{V^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})}} \le t \mid w^S\right\} - \Phi(t)\right| \\
&= \sup_{t\in\mathbb{R}}\left| P\left\{\frac{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) - \tau}{\sqrt{v_I(w^S)}} \le t \mid w^S\right\} - \Phi\left(t\frac{\sqrt{v_I(w^S) + S_\tau^2(w^S)/I}}{\sqrt{V^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})}}\right)\right| \\
&\le \sup_{t\in\mathbb{R}}\left| P\left\{\frac{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) - \tau}{\sqrt{v_I(w^S)}} \le t \mid w^S\right\} - \Phi(t)\right| + \sup_{t\in\mathbb{R}}\left| \Phi\left(t\frac{\sqrt{v_I(w^S) + S_\tau^2(w^S)/I}}{\sqrt{V^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})}}\right) - \Phi(t)\right| \qquad (10)
\end{aligned}$$

In Lemma B.1, we show almost sure convergence of $V^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})I$ to $v_I(w^S)I + S^2_\tau(w^S)$, which implies that, as $I, J \to \infty$,

$$\sup_{t\in\mathbb{R}} \left| \Phi\left(t\frac{\sqrt{v_I(w^S) + S^2_\tau(w^S)/I}}{\sqrt{V^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})}} \right) - \Phi(t) \right| \xrightarrow{a.s.} 0 \tag{11}$$

Then we define the average treatment effect $\tau(w^S)$ on the focal units:

$$\tau(w^S) = \frac{1}{IJ_0}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{i,j}(1,0)(1-w_j^S) - \frac{1}{IJ_0}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{i,j}(0,0)(1-w_j^S) = \frac{1}{I}\sum_{i=1}^{I}\bar{Y}_i^B(1) - \frac{1}{I}\sum_{i=1}^{I}\bar{Y}_i^B(0)\ .$$

Following Masoero et al. (2024), define $\mathcal{E}_{1,\eta}$ and $\mathcal{E}_{2,\eta}$ as follows:

$$\begin{aligned}
\mathcal{E}_{1,\eta} &:= \left\{ w^S : \left| v_I(w^S) - \mathrm{Var}\{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})\} \right| \le \Delta_{1,\eta} \right\} \\
\mathcal{E}_{2,\eta} &:= \left\{ w^S : |\tau(w^S) - \tau| \le \Delta_{2,\eta} \right\}\ ,
\end{aligned}$$

where

$$\begin{aligned}
\Delta_{1,\eta} &:= CC_2^2\left[\Delta_I\Delta_J + \Delta_J^2\right]\log(1/\eta) \\
\Delta_{2,\eta} &:= CC_2\Delta_J\log(C'/\eta) \\
\Delta_I &:= I_0^{-1} + I_1^{-1} \\
\Delta_J &:= J_0^{-1} + J_1^{-1}\ .
\end{aligned}$$

Note that

$$\begin{aligned}
&\sup_{t\in\mathbb{R}} \left| P\left\{ \frac{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) - \tau}{\sqrt{v_I(w^S)}} \le t \mid w^S \right\} - \Phi(t) \right| \\
&= \sup_{t\in\mathbb{R}} \left| P\left\{ T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) \le t \mid w^S \right\} - \Phi\left(\frac{t-\tau}{\sqrt{v_I(w^S)}} \right) \right| \\
&\le \sup_{t\in\mathbb{R}} \left| P\left\{ T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) \le t \mid w^S \right\} - \Phi\left(\frac{t-\tau(w^S)}{\sqrt{v_I(w^S)}} \right) \right| + \sup_{t\in\mathbb{R}} \left| \Phi\left(\frac{t-\tau(w^S)}{\sqrt{v_I(w^S)}} \right) - \Phi\left(\frac{t-\tau}{\sqrt{v_I(w^S)}} \right) \right| \\
&\le \sup_{t\in\mathbb{R}} \left| P\left\{ \frac{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) - \tau(w^S)}{\sqrt{v_I(w^S)}} \le t \mid w^S \right\} - \Phi(t) \right| + \sup_{t\in\mathbb{R}} \left| \Phi\left(\frac{t-\tau(w^S)}{\sqrt{v_I(w^S)}} \right) - \Phi\left(\frac{t-\tau}{\sqrt{v_I(w^S)}} \right) \right|\ .
\end{aligned}$$

By Lemma B.5 of Masoero et al. (2024), given $w^S \in \mathcal{E}_{1,\eta}$, we have

$$\sup_{t\in\mathbb{R}} \left| P\left\{ \frac{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) - \tau(w^S)}{\sqrt{v_I(w^S)}} \le t \mid w^S \right\} - \Phi(t) \right| \le \frac{C\left(\Delta_I C_2 + \sqrt{\Delta_{1,\eta}}\right)}{\mathrm{Var}\{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})\}^{\frac{1}{2}}},$$

Next, note that

$$\begin{aligned}
\sup_{t\in\mathbb{R}} \left| \Phi\left(\frac{t-\tau(w^S)}{\sqrt{v_I(w^S)}} \right) - \Phi\left(\frac{t-\tau}{\sqrt{v_I(w^S)}} \right) \right| &\le \sup_{t\in\mathbb{R}} \left| \Phi\left(\frac{t-\tau(w^S)}{\sqrt{v_I(w^S)}} \right) - \Phi\left(\frac{t-\tau(w^S)}{\mathrm{Var}\{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})\}^{\frac{1}{2}}} \right) \right| \\
&\quad + \sup_{t\in\mathbb{R}} \left| \Phi\left(\frac{t-\tau}{\sqrt{v_I(w^S)}} \right) - \Phi\left(\frac{t-\tau(w^S)}{\mathrm{Var}\{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})\}^{\frac{1}{2}}} \right) \right|
\end{aligned}$$

Following Lemma B.17 in Masoero et al. (2024), for any $\eta \le 1$, we have

$$\sup_{t\in\mathbb{R}}\left|\Phi\left(\frac{t-\tau(w^S)}{\sqrt{v_I(w^S)}}\right)-\Phi\left(\frac{t-\tau(w^S)}{\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}^{\frac{1}{2}}}\right)\right|$$
$$\le \eta+\frac{2|\sqrt{v_I(w^S)}-\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}^{\frac{1}{2}}|\sqrt{\log(e/\eta)}}{\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}^{\frac{1}{2}}}$$

and

$$\sup_{t\in\mathbb{R}}\left|\Phi\left(\frac{t-\tau}{\sqrt{v_I(w^S)}}\right)-\Phi\left(\frac{t-\tau(w^S)}{\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}^{\frac{1}{2}}}\right)\right|$$
$$\le \eta+\frac{|\tau-\tau(w^S)|+2|\sqrt{v_I(w^S)}-\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}^{\frac{1}{2}}|\sqrt{\log(e/\eta)}}{\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}^{\frac{1}{2}}}\ .$$

Note that, under $w^S \in \mathcal{E}_{1,\eta}$,

$$\left|\sqrt{v_I(w^S)}-\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}^{\frac{1}{2}}\right|\le\sqrt{|v_I(w^S)-\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}|}\le\sqrt{\Delta_{1,\eta}}\ .$$

Therefore, under $w^S \in \mathcal{E}_{1,\eta}\bigcap\mathcal{E}_{2,\eta}$ we have

$$\sup_{t\in\mathbb{R}}\left|\Phi\left(\frac{t-\tau(w^S)}{\sqrt{v_I(w^S)}}\right)-\Phi\left(\frac{t-\tau}{\sqrt{v_I(w^S)}}\right)\right|\le 2\eta+\frac{\Delta_{2,\eta}+4\sqrt{\Delta_{1,\eta}\log(e/\eta)}}{\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}^{\frac{1}{2}}}\ .$$

Denote by $\Delta'(\eta)$ the summation of the two upper bounds:

$$\Delta'(\eta)=2\eta+\frac{C\left(\Delta_I C_2+\sqrt{\Delta_{1,\eta}}\right)}{\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}^{\frac{1}{2}}}+\frac{\Delta_{2,\eta}+4\sqrt{\Delta_{1,\eta}\log(e/\eta)}}{\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}^{\frac{1}{2}}}\ .$$

Note that $\Delta'(\eta)$ is invariant to the seller side treatment allocation w^S. Next, consider

$$\begin{aligned}\mathbb{E}\sup_{t\in\mathbb{R}}\left|P\left\{\frac{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})-\tau}{\sqrt{v_I(w^S)}}\le t\mid w^S\right\}-\Phi(t)\right| &\le \mathbb{E}\left[1\{w^S\in\mathcal{E}_{1,\eta}\cap\mathcal{E}_{2,\eta}\}\Delta'(\eta)+1\{w^S\notin\mathcal{E}_{1,\eta}\cap\mathcal{E}_{2,\eta}\}\right]\\ &\le \Delta'(\eta)+\left[1-P(w^S\in\mathcal{E}_{1,\eta}\cap\mathcal{E}_{2,\eta})\right]\ .\end{aligned}$$

Following Corollary B.21 by Masoero et al. (2024), there exists a universal constant $C>0$ such that

$$\mathbb{E}\sup_{t\in\mathbb{R}}\left|P\left\{\frac{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})-\tau}{\sqrt{v_I(w^S)}}\le t\mid w^S\right\}-\Phi(t)\right|\le 2\eta+\Delta'(\eta)\le C\Delta\log\left(\frac{C}{\Delta}\right)\ , \tag{12}$$

where

$$\Delta:=\frac{C_1C_2(I^{-1}+J^{-1})}{\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}^{\frac{1}{2}}}\ .$$

Under $H_0^{wb}:\tau=0$, we have

$$\sup_{t\in\mathbb{R}}\left|P\left\{\frac{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})}{\sqrt{V^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})}}\sqrt{\frac{v_I(w^S)+S_\tau^2(w^S)/I}{v_I(w^S)}}\le t\right\}-\Phi(t)\right|$$

$$
\begin{aligned}
&= \sup_{t\in\mathbb{R}} \left| \mathbb{E}\left[P\left\{ \frac{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})}{\sqrt{V^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})}} \sqrt{\frac{v_I(w^S)+S_\tau^2(w^S)/I}{v_I(w^S)}} \le t \mid w^S \right\} - \Phi(t) \right] \right| \\
&\le \mathbb{E}\sup_{t\in\mathbb{R}} \left| P\left\{ \frac{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})}{\sqrt{v_I(w^S)}} \le t \mid w^S \right\} - \Phi(t) \right| + \mathbb{E}\sup_{t\in\mathbb{R}} \left| \Phi\left(t\frac{\sqrt{v_I(w^S)+S_\tau^2(w^S)/I}}{\sqrt{V^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})}} \right) - \Phi(t) \right| \\
&\le C\Delta \log\left(\frac{C}{\Delta}\right) + \mathbb{E}\sup_{t\in\mathbb{R}} \left| \Phi\left(t\frac{\sqrt{v_I(w^S)+S_\tau^2(w^S)/I}}{\sqrt{V^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})}} \right) - \Phi(t) \right| .
\end{aligned}
$$

The first inequality follows by Jensen's inequality and (10). The second inequality follows (12). Then, by dominated convergence theorem and (11), we have

$$
\lim_{I,J\to\infty} \mathbb{E}\sup_{t\in\mathbb{R}} \left| \Phi\left(t\frac{\sqrt{v_I(w^S)+S_\tau^2(w^S)/I}}{\sqrt{V^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})}} \right) - \Phi(t) \right| = 0 .
$$

To interpret the limit of Δ, we follow Masoero et al. (2024) by noting that $\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}$ will be of order $(I^{-1}+J^{-1})$. Therefore, we have

$$
\lim_{I,J\to\infty} C\Delta \log\left(\frac{C}{\Delta}\right) = 0 .
$$

Finally, since $(v_I(w^S)+S_\tau^2(w^S)/I)/v_I(w^S) \ge 1$, $T^{WB}(\mathbf{W}\mid\mathbf{Y},\mathcal{C})$ is stochastically dominated by $N(0,1)$.

B.2.2 Randomization Distribution

Next, we show that the randomization distribution of $T^{WB}(\mathbf{W}\mid\mathbf{Y},\mathcal{C})$ converges to $N(0,1)$ in distribution. We denote the randomization distribution of the test statistic and variance estimator with $T^{B,\pi}(\mathbf{W}\mid\mathbf{Y},\mathcal{C})$ and $V^{B,\pi}(\mathbf{W}\mid\mathbf{Y},\mathcal{C})$. Our conditional randomization test takes $\{\bar{Y}_i^B\}_{i=1}^I$ as the fixed input for permuting the treatment vector, and thereby induces a sequence of finite populations $\{\bar{Y}_i^B, \bar{Y}_i^B\}_{i=1}^I$ as the pseudo potential outcomes under both treatment and control. Therefore, the proof should follow exactly Appendix B.2.1. Given a fixed w^B, with a slight abuse of notation, we should have the same Berry-Esseen type result from Lemma B.5 of Masoero et al. (2024), under $w^S \in \mathcal{E}_{1,\eta}$,

$$
\sup_{t\in\mathbb{R}} \left| P\left\{ \frac{T^{B,\pi}(\mathbf{W}\mid\mathbf{Y},\mathcal{C})-\tau(w^S)}{\sqrt{v_I(w^S)}} \le t \mid w^S \right\} - \Phi(t) \right| \le \frac{C\left(\Delta_I C_2 + \sqrt{\Delta_{1,\eta}}\right)}{\mathrm{Var}\{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})\}^{\frac{1}{2}}},
$$

where $\tau(w^S)=0$ and $v_I(w^S) = \frac{I}{I_1 I_0} S^2(w^S)$ with $S^2(w^S) = \frac{1}{I-1}\sum_{i=1}^N (\bar{Y}_i^B - \bar{Y}^B)^2$ and $\bar{Y}^B = \frac{1}{I-1}\sum_{i=1}^N \bar{Y}_i^B$. Note that, under $w^S \in \mathcal{E}_{1,\eta}$,

$$
\begin{aligned}
&\sup_{t\in\mathbb{R}} \left| P\left\{ \frac{T^{B,\pi}(\mathbf{W}\mid\mathbf{Y},\mathcal{C})}{\sqrt{V^{B,\pi}(\mathbf{W}\mid\mathbf{Y},\mathcal{C})}} \le t \mid w^S \right\} - \Phi(t) \right| \\
&\le \sup_{t\in\mathbb{R}} \left| P\left\{ \frac{T^B(\mathbf{W}\mid\mathbf{Y},\mathcal{C})}{\sqrt{v_I(w^S)}} \le t \mid w^S \right\} - \Phi(t) \right| + \sup_{t\in\mathbb{R}} \left| \Phi\left(t\sqrt{\frac{v_I(w^S)}{V^{B,\pi}(\mathbf{W}\mid\mathbf{Y},\mathcal{C})}} \right) - \Phi(t) \right| .
\end{aligned}
$$

The results in Lemma B.1 can be adapted to obtain $V^{B,\pi}(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) \xrightarrow{a.s.} v_I(w^S)$ as $I, J \to \infty$. Then, the rest of the proof should be equivalent to Appendix B.2.1 to finally obtain

$$\lim_{I,J\to\infty} \sup_{t\in\mathbb{R}} \left| P\left\{ \frac{T^{B,\pi}(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})}{\sqrt{V^{B,\pi}(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})}} \leq t \right\} - \Phi(t) \right| = 0 \ .$$

■

B.2.3 Discussions on Total Effects

The proof of validity for total effect tests under the weak null follows the same strategy; thus, we only present the key step. Without loss of generality, we consider a 1-block conditioning event consisting of diagonal elements from the set of focal units $\{(i,i) : w_i^B = w_i^S\}$ from a square matrix, i.e., $I = J = n$. Then, the conditional randomization space is defined as $\mathcal{W} = \{\mathbf{W} : w^B = w^S\}$. The key step in the proof requires showing that

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} \left| P\left\{ \frac{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) - \tau}{\sqrt{v_I(\mathcal{W})}} \leq t \right\} - \Phi(t) \right| = 0 \ , \tag{13}$$

where $v_I(\mathcal{W})$ is defined similarly to $V_I(w^S)$. Consider the conditional probability given as follows:

$$\begin{aligned} &\sup_{t\in\mathbb{R}} \left| P\left\{ \frac{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) - \tau}{\sqrt{v_I(\mathcal{W})}} \leq t \mid \mathcal{W} \right\} - \Phi(t) \right| \\ &\leq \sup_{t\in\mathbb{R}} \left| P\left\{ \frac{T^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C}) - \tau(\mathcal{W})}{\sqrt{v_I(\mathcal{W})}} \leq t \mid \mathcal{W} \right\} - \Phi(t) \right| + \sup_{t\in\mathbb{R}} \left| \Phi\left(t - \frac{\tau(\mathcal{W}) - \tau}{\sqrt{v_I(\mathcal{W})}} \right) - \Phi(t) \right| \ . \end{aligned}$$

By Theorem 1 of Li and Ding (2017), as $n \to \infty$, the first term converges to 0 under any $\mathcal{W}$. For the second term, by the mean value theorem, there exists c such that

$$\sup_{t\in\mathbb{R}} \left| \Phi\left(t - \frac{\tau(\mathcal{W}) - \tau}{\sqrt{v_I(\mathcal{W})}} \right) - \Phi(t) \right| \leq \phi(c) \left| \frac{\tau(\mathcal{W}) - \tau}{\sqrt{v_I(\mathcal{W})}} \right| \leq \frac{1}{\sqrt{2\pi}} \left| \frac{\tau(\mathcal{W}) - \tau}{\sqrt{v_I(\mathcal{W})}} \right| \ .$$

The key is to show $\lim_{n\to\infty} E[|\tau(\mathcal{W}) - \tau|] = 0$, which follows by Assumption (b) that $\tau(\mathcal{W})$ is almost surely bounded, and $|\tau(\mathcal{W}) - \tau| \xrightarrow{p} 0$. The latter is equivalent to showing that the average of a random subset of n observations from n^2 samples converges to the average over n^2 samples.[4] By Lemma B.2, the convergence in probability holds. Therefore, equation (13) holds.

B.2.4 Auxiliary Lemmas

Lemma B.1. *Suppose that assumptions in Theorem 4.1 hold, conditional on seller side randomization w^S, as $I, J \to \infty$, $V^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})I \xrightarrow{a.s.} v_I(w^S)I + S_\tau^2(w^S)$.*

[4] Though the subsampling step is subject to the constraint that there is only one treated pair in each row and column, the total number of restrictions ($Z_{i,1}Z_{i,j} = 0$ and $Z_{1,j}Z_{i,j} = 0$ for all $i \neq j$) is equal to $2n(n-1)$, which is of lower order than n^4. Therefore, when n is large, $\mathrm{Cov}(Z_i, Z_j)$ would converge to what we obtained in Lemma B.2.

Proof. Our goal is to show that

$$\frac{I}{I_1}\frac{1}{I_1-1}\sum_{i=1}^{I} w_i^B\left(\bar{Y}_i^B-\hat{Y}(1)\right)^2+\frac{I}{I_0}\frac{1}{I_0-1}\sum_{i=1}^{I}(1-w_i^B)\left(\bar{Y}_i^B-\hat{Y}(0)\right)^2 \xrightarrow{a.s.} \frac{I}{I_1}S_1^2(w^S)+\frac{I}{I_0}S_0^2(w^S)\ .$$

It is sufficient to show that

$$\frac{1}{I_1-1}\sum_{i=1}^{I} w_i^B\left(\bar{Y}_i^B(1)-\hat{Y}(1)\right)^2 \xrightarrow{a.s.} S_1^2(w^S)\ .$$

First, define $\tilde{S}_z^2(w^S)=(I_1-1)^{-1}\sum_{i=1}^{I} w_i^B(\bar{Y}_i^B(z)-\bar{Y}_\cdot^B(z))^2$. Note that we have

$$S_z^2(w^S)=\frac{1}{I-1}\sum_{i=1}^{I}(\bar{Y}_i^B(z)-\bar{Y}_\cdot^B(z))^2 \le \frac{2}{I-1}\sum_{i=1}^{I}\bar{Y}_i^B(z)^2+\bar{Y}_\cdot^B(z)^2 \le 4C_2^2 \tag{14}$$

and thus

$$\frac{1}{I}\sum_{i=1}^{I}(\bar{Y}_i^B(z)-\bar{Y}_\cdot^B(z))^4 \le \frac{1}{I}\sum_{i=1}^{I}(\bar{Y}_i^B(z)^2+\bar{Y}_\cdot^B(z)^2)^2 \le 4C_2^4\ .$$

By following the proof of Lemma A3 in Wu and Ding (2021), as $I, J \to \infty$,

$$\begin{aligned}\left|\frac{1}{I_1-1}\sum_{i=1}^{I} w_i^B\left(\bar{Y}_i^B(1)-\hat{Y}(1)\right)^2-\tilde{S}_1^2(w^S)\right| &= \left|\frac{1}{I_1-1}\sum_{i=1}^{I} w_i^B\left(\left(\bar{Y}_i^B(1)-\hat{Y}(1)\right)^2-\left(\bar{Y}_i^B(1)-\bar{Y}_\cdot^B(1)\right)^2\right)\right| \\ &= \left|\frac{-I_1}{I_1-1}(\hat{Y}(1)-\bar{Y}_\cdot^B(1))^2\right| \xrightarrow{a.s.} 0\ .\end{aligned}$$

Then, as $I, J \to \infty$,

$$\begin{aligned}&\left|\tilde{S}_z^2(w^S)-S_z^2(w^S)\right| \\ &\le \frac{I_1}{I_1-1}\left|\frac{1}{I_1}\sum_{i=1}^{I} w_i^B\left(\bar{Y}_i^B(1)-\bar{Y}_\cdot^B(1)\right)^2-\frac{1}{I}\sum_{i=1}^{I}\left(\bar{Y}_i^B(1)-\bar{Y}_\cdot^B(1)\right)^2\right|+\frac{1}{I_1-1}2C_2^2 \xrightarrow{a.s.} 0\ .\end{aligned}$$

Therefore,

$$\begin{aligned}&\left|\frac{1}{I_1-1}\sum_{i=1}^{I} w_i^B\left(\bar{Y}_i^B(1)-\hat{Y}(1)\right)^2-S_1^2(w^S)\right| \\ &\le \left|\frac{1}{I_1-1}\sum_{i=1}^{I} w_i^B\left(\bar{Y}_i^B(1)-\hat{Y}(1)\right)^2-\tilde{S}_1^2(w^S)\right|+\left|\tilde{S}_z^2(w^S)-S_z^2(w^S)\right| \xrightarrow{a.s.} 0\ .\end{aligned}$$

We now finally have $V^B(\mathbf{W} \mid \mathbf{Y}, \mathcal{C})I \xrightarrow{a.s.} v_I(w^S)+S_\tau^2(w^S)$. ■

Lemma B.2. *Let $\{X_i\}_{i=1}^{n^2}$ be a collection of real-valued non-random population quantities. Consider drawing a random sample of size n. Let $\{Z_i\}_{i=1}^{n^2}$ be a sequence of binary sampling indicators such that $Z_i = 1$ if i is* *sampled. If there exist $C > 0$ such that $|X_i| \le C$ for all $1 \le i \le n^2$, as $n \to \infty$,*

$$\left| \frac{1}{n} \sum_{i=1}^{n^2} X_i Z_i - \frac{1}{n^2} \sum_{i=1}^{n^2} X_i \right| \xrightarrow{p} 0 \ .$$

Proof. Note that $E[Z_i] = 1/n$ and $\mathrm{Var}(Z_i) = 1/n - 1/n^2$. And

$$\mathrm{Var}\left(\sum_{i=1}^{n^2} Z_i \right) = n^2 \,\mathrm{Var}(Z_i) + n^2(n^2 - 1)\,\mathrm{Cov}(Z_i, Z_j) = 0 = O(1/n^4) \ ,$$

which implies $\mathrm{Cov}(Z_i, Z_j) = -\frac{\mathrm{Var}(Z_i)}{n^2-1} = -\frac{1}{n(n^2-1)}\left(1 - \frac{1}{n}\right)$.

$$E\left[\frac{1}{n} \sum_{i=1}^{n^2} X_i Z_i - \frac{1}{n^2} \sum_{i=1}^{n^2} X_i \right] = 0 \ .$$

By Chebyshev's inequality,

$$\begin{aligned} P\left(\left| \frac{1}{n} \sum_{i=1}^{n^2} X_i Z_i - \frac{1}{n^2} \sum_{i=1}^{n^2} X_i \right| \ge \epsilon \right) &\le \frac{\mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n^2} X_i Z_i \right)}{\epsilon} \\ &= \frac{1}{\epsilon n^2} \left(\mathrm{Var}(Z_i) \sum_{i=1}^{n^2} X_i^2 + \mathrm{Cov}(Z_i, Z_j) \sum_{i=1}^{n^2} \sum_{j \ne i} X_i X_j \right) \\ &= \frac{\mathrm{Var}(Z_i)}{\epsilon n^2} \left(\sum_{i=1}^{n^2} X_i^2 - \frac{1}{n^2 - 1} \sum_{i=1}^{n^2} \sum_{j \ne i} X_i X_j \right) \\ &= \frac{\mathrm{Var}(Z_i)}{\epsilon n^2} \frac{1}{n^2 - 1} \left(n^2 \sum_{i=1}^{n^2} X_i^2 - \left(\sum_{i=1}^{n^2} X_i \right)^2 \right) \\ &\le \frac{\mathrm{Var}(Z_i)}{\epsilon n^2} \frac{1}{n^2 - 1} n^4 C = \left(\frac{1}{n} - \frac{1}{n^2} \right) \frac{n^4 C}{\epsilon n^2 (n^2 - 1)} \to 0 \ . \end{aligned}$$

Therefore, the convergence in probability holds. ■

References

Abadie, A., Athey, S., Imbens, G. W. and Wooldridge, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, **88** 265–296.

Aronow, P. M. (2012). A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research*, **41** 3–16.

Athey, S., Eckles, D. and Imbens, G. W. (2018). Exact p-values for network interference. *Journal of the American Statistical Association*, **113** 230–240.

Athey, S. and Luca, M. (2019). Economists (and economics) in tech companies. *Journal of Economic Perspectives*, **33** 209–30.

Bajari, P., Burdick, B., Imbens, G. W., Masoero, L., McQueen, J., Richardson, T. S. and Rosen, I. M. (2023). Experimental Design in Marketplaces. *Statistical Science*, **38** 458 – 476.

Basse, G., Ding, P., Feller, A. and Toulis, P. (2024). Randomization tests for peer effects in group formation experiments. *Econometrica*, **92** 567–590.

Basse, G. W., Feller, A. and Toulis, P. (2019). Randomization tests of causal effects under interference. *Biometrika*, **106** 487–494.

Blake, T. and Coey, D. (2014). Why marketplace experimentation is harder than it seems: the role of test-control interference. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*. EC '14, Association for Computing Machinery, New York, NY, USA, 567–582. URL https://doi.org/10.1145/2600057.2602837.

Canay, I. A., Romano, J. P. and Shaikh, A. M. (2017). Randomization tests under an approximate symmetry assumption. *Econometrica*, **85** 1013–1030.

Chung, E. and Romano, J. P. (2013). Exact and asymptotically robust permutation tests. *The Annals of Statistics*, **41** 484 – 507.

Comola, M. and Prina, S. (2021). Treatment effect accounting for network changes. *The Review of Economics and Statistics*, **103** 597–604.

DiCiccio, C. J. and Romano, J. P. (2017a). Robust permutation tests for correlation and regression coefficients. *Journal of the American Statistical Association*, **112** 1211–1220.

DiCiccio, C. J. and Romano, J. P. (2017b). Robust permutation tests for correlation and regression coefficients. *Journal of the American Statistical Association*, **112** 1211–1220.

Dupas, P., Green, S., Keats, A. and Robinson, J. (2016). 63challenges in banking the rural poor: Evidence from kenya's western province. In *African Successes, Volume III: Modernization and Development*. University of Chicago Press. https://academic.oup.com/chicago-scholarship-online/book/0/chapter/183709565/chapter-ag-pdf/44675132/book_22957_section_183709565.ag.pdf.

Fisher, R. A. (1953). *The design of experiments*. Oliver & Boyd, Edinburgh.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press.

Johari, R., Li, H., Liskovich, I. and Weintraub, G. Y. (2022). Experimental design in two-sided platforms: An analysis of bias. *Management Science*, **68** 7069–7089.

Kohavi, R., Tang, D. and Xu, Y. (2020). *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*. Cambridge University Press.

Li, X. and Ding, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, **112** 1759–1769.

Masoero, L., Vijaykumar, S., Richardson, T., McQueen, J., Rosen, I., Burdick, B., Bajari, P. and Imbens, G. (2024). Multiple randomization designs: Estimation and inference with interference. 2401.01264.

Prina, S. (2015). Banking the poor via savings accounts: Evidence from a field experiment. *Journal of Development Economics*, **115** 16–31.

Puelz, D., Basse, G., Feller, A. and Toulis, P. (2021). A Graph-Theoretic Approach to Randomization Tests of Causal Effects under General Interference. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **84** 174–204.

Rubin, D. (1972). Estimating causal effects of treatments in experimental and observational studies. *ETS Research Bulletin Series*, **1972** i–31.

Rubin, D. B. (1980). Comment. *Journal of the American Statistical Association*, **75** 591–593.

Thomke, S. (2003). *Experimentation Matters: Unlocking the Potential of New Technologies for Innovation*. Harvard Business School Press, Boston, MA.

Toulis, P. and Parkes, D. C. (2016). Long-term causal effects via behavioral game theory. *Advances in Neural Information Processing Systems*, **29**.

Wager, S. and Xu, K. (2021). Experimenting in equilibrium. *Management Science*, **67** 6694–6715.

Wu, J. and Ding, P. (2021). Randomization tests for weak null hypotheses in randomized experiments. *Journal of the American Statistical Association*, **116** 1898–1913.

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