

Cluster-robust inference with a single treated cluster using the t -test*

Chun Pong Lau[†] Xinran Li[‡]

November 7, 2025

Abstract

This paper considers inference when there is a single treated cluster and a fixed number of control clusters, a setting that is common in empirical work, especially in difference-in-differences designs. We use the t -statistic and develop suitable critical values to conduct valid inference under weak assumptions allowing for unknown dependence within clusters. In particular, our inference procedure does not involve variance estimation. It only requires specifying the relative heterogeneity between the variances from the treated cluster and some, but not necessarily all, control clusters. Our proposed test works for any significance level when there are at least two control clusters. When the variance of the treated cluster is bounded by those of all control clusters up to some prespecified scaling factor, the critical values for our t -statistic can be easily computed without any optimization for many conventional significance levels and numbers of clusters. In other cases, one-dimensional numerical optimization is needed and is often computationally efficient. We have also tabulated common critical values in the paper so researchers can use our test readily. We illustrate our method in simulations and empirical applications.

Keywords: Cluster-robust inference, single treated cluster, t -test, simultaneous inference, difference-in-differences.

*We thank helpful comments from Stéphane Bonhomme, Ivan Canay, Azeem Shaikh, and Alex Torgovitsky.

[†]Kenneth C. Griffin Department of Economics, The University of Chicago, ccplau@uchicago.edu.

[‡]Department of Statistics, The University of Chicago, xinranli@uchicago.edu.

1 Introduction

In difference-in-differences designs, it is common for researchers to conduct inference using cluster-robust methods to account for correlation within clusters. However, inference becomes challenging when there is only a single treated cluster. Intuitively, this is because researchers are faced with only one estimate from the treated cluster, making its uncertainty difficult to quantify. Existing methods either impose strong assumptions on having a large number of clusters, require the variances to be homogeneous or estimable, or only work for specific significance levels depending on the number of clusters. In this paper, we develop a t -test associated with a suitable critical value to conduct valid inference that relaxes these assumptions.

Having a single treated cluster and a finite number of control clusters is common in empirical work. For instance, this can happen when researchers use data in the United States to examine the impact of a policy that takes place in one particular state but not the other states. The nearby states or the remaining states are commonly used as the control group. In these scenarios, it is reasonable for researchers to believe they are not in the scenario with a “large” number of control clusters. Some recent examples in this context include [Wang and Burke \(2022\)](#) on the effect of payday loan regulations in Texas, [Harris and Larsen \(2023\)](#) on the effect of Hurricane Katrina on student outcomes in New Orleans, [Dillender et al. \(2023\)](#) on the effect of change in health care reimbursement rates in Illinois, [Alpert et al. \(2024\)](#) on the impact of Kentucky’s prescription drug monitoring programs on opioid prescribing, and [Kumar and Liang \(2024\)](#) on the labor market effects of constitutional amendments in Texas. [Hagemann \(2024\)](#) has documented earlier related examples in the context of a single treated cluster. In the next section of this paper, we demonstrate that our test is also applicable to other empirical designs, in addition to difference-in-differences.

This paper contributes to the literature on cluster-robust inference. We assume that the number of clusters is fixed, as in some related work in this literature such as [Bester et al. \(2011\)](#), [Ibragimov and Müller \(2010, 2016\)](#), [Canay et al. \(2017\)](#), [Hagemann \(2022\)](#), and [Lau \(2025\)](#). Although the tests in the aforementioned papers are valid when there is a finite number of clusters, they are not suitable for the problem with a single treated cluster. The first test requires certain homogeneity conditions, and the other tests cannot be applied when there is a single treated cluster. [Canay et al. \(2021\)](#) has showed that Wild cluster bootstrap popularized by [Cameron et al. \(2008\)](#) can be valid under strong homogeneity conditions when there is a fixed number of clusters. See, for instance, [Cameron and Miller \(2015\)](#), [Conley et al. \(2018\)](#), [MacKinnon et al. \(2023\)](#), and [Alvarez et al. \(2025\)](#) for some

surveys on the literature of conducting inference with a fixed number of clusters.

Several tests have been developed to conduct inference when there is a single treated cluster, but with assumptions that can be strong in practice. [Conley and Taber \(2011\)](#) assume homogeneous errors. [Ferman and Pinto \(2019\)](#) relax the homogeneity assumption and allow for known heteroskedasticity. [Alvarez and Ferman \(2023\)](#) relax the homogeneity assumption and allows for spatial correlation. All these tests assume that there is an infinite number of control clusters. Our t -test neither assumes the variances are known nor assumes an infinite number of control clusters. Recently, [Hagemann \(2024\)](#) develops a novel rearrangement test that relaxes the assumptions mentioned earlier. [Hagemann \(2024\)](#) is the most related paper in that his test relies on a relative heterogeneity condition between the treated and control clusters. Both our test and [Hagemann \(2024\)](#) work with a fixed number of clusters, allow for arbitrary correlation within clusters, and do not assume that we can estimate or know the cluster variances. He shows his test is valid when the relative heterogeneity condition holds with all or all but one control cluster. However, the validity of his test depends on the number of clusters, the significance level, and the relative heterogeneity condition. For example, when the standard deviation of the treated cluster is bounded by 2 times all but one of the standard deviations of the control clusters, [Hagemann \(2024\)](#) requires at least 14 control clusters to conduct a one-sided test at the 2.5% level using his rearrangement test, and at least 17 control clusters are needed for a 1% level test¹. This potentially limits the applicability of his test. Among the cases where his test is valid, he computes weights for his rearrangement test using numerical optimization that lead to valid tests.

There are several important distinctions between our work and [Hagemann \(2024\)](#). First, we show that our test can be valid for any choice of significance levels and heterogeneity parameters when there are at least two control clusters. For many conventional combinations of significance levels and number of clusters, we derive a closed-form critical value for the t -test. No optimization is necessary for these cases, so researchers can apply the test readily. For other cases, we can derive critical values that lead to valid tests through one-dimensional optimizations. Therefore, our test can have broader applicability. Second, we relax the relative heterogeneity condition in [Hagemann \(2024\)](#). Intuitively, the condition restricts the relative variance between the treated and control clusters. [Hagemann \(2024\)](#) requires such condition to hold between the treated and all (or all but one) control clusters to show the validity of his test. We weaken this condition by allowing the researcher to impose this restriction between the treated cluster and

¹Equivalently, to conduct two-sided tests, there have to be at least 14 control clusters for a 5% level test.

any number of control clusters. This allows researchers to bound relative heterogeneity on variances depending on their choice on how to restrict the amount of heterogeneity between the treated and control clusters. Third, we allow for simultaneous inference across all the relative heterogeneity assumptions that we impose. Specifically, under the assumption of no treatment effect, we can infer the minimum amount of relative heterogeneity required to explain away the observed association between treatment and outcomes. If the true relative heterogeneity is unlikely to exceed this amount, the treatment is likely to have a significant nonzero effect.

While the t -test has been used in [Bakirov and Székely \(2006\)](#) and [Ibragimov and Müller \(2010, 2016\)](#), our proof on the validity of the t -test in the current context requires different proof strategies. This is because we only have a single treated cluster and the relative heterogeneity assumption imposes additional structure on the parameter space.

We confirm the performance of our proposed test in simulations. We find that our test controls size under a wide range of significance levels and number of clusters. We also have favorable power performance when compared to other tests in the literature.

The rest of the paper is organized as follows. Section 2 outlines some popular empirical designs that our test can be applied to. Section 3 presents our inference procedure. Readers interested in applying the test can refer to Algorithm 3.1. Section 4 presents our main theory. Section 5 explains how our test can be easily used for simultaneous inference. Section 6 presents the simulation results. Section 7 contains two empirical studies. Section 8 concludes. All proofs can be found in the supplementary material.

2 Motivating examples

We start by describing several empirically relevant designs that are common in applied work and are related to the issue of having a single treated cluster. These examples show that our method applies more broadly apart from standard difference-in-differences designs. The first three examples have also been discussed in [Hagemann \(2024\)](#).

In the following, $i \in \mathcal{I} \equiv \{1, \dots, n\}$ indices units, $j \in \mathcal{J} \equiv \{1, \dots, m+1\}$ indices clusters, and $t \in \mathcal{T} \equiv \{1, \dots, T\}$ indices time. Only cluster $(m+1)$ is treated, and the remaining clusters are controls. Thus, \mathcal{J} can be partitioned as $\mathcal{J}_0 \cup \mathcal{J}_1$, where $\mathcal{J}_0 \equiv \{1, \dots, m\}$ and $\mathcal{J}_1 \equiv \{m+1\}$ denote the control clusters and treated cluster respectively. Let $D_j \equiv \mathbb{1}[j \in \mathcal{J}_1]$ be a cluster-level treatment indicator across $j \in \mathcal{J}$.

As to be described in Section 3, we require researchers to be able to run regression

cluster by cluster, and write the estimator for the target parameter Δ in terms of the estimator for parameters from cluster-level regressions, denoted by $\{\hat{\theta}_j\}_{j=1}^{m+1}$. The following examples explain how to connect $\{\hat{\theta}_j\}_{j=1}^{m+1}$ with the estimator for the target parameters in common empirical designs.

Example 2.1 (Clustered regression). Consider the following model

$$Y_{ij} = \beta_0 + \Delta D_j + U_{ij},$$

where $\beta_0, \Delta, U_{ij} \in \mathbb{R}$. Under the assumption that $\mathbb{E}[U_{ij}|D_j] = 0$, Δ can be estimated by

$$\hat{\Delta} = \hat{\theta}_{m+1} - \frac{1}{m} \sum_{j=1}^m \hat{\theta}_j,$$

where $\hat{\theta}_j$ is the sample mean of Y_{ij} for each $j \in \mathcal{J}$. △

Example 2.2 (Difference-in-differences). Suppose $T = 2$ and let $\text{Post}_t \equiv \mathbb{1}[t = 2]$ for all $t \in \mathcal{T}$. Consider the following model

$$Y_{jt} = \alpha_j + \beta \text{Post}_t + \Delta D_j \text{Post}_t + U_{jt}, \quad (1)$$

where $U_{jt} \in \mathbb{R}$, and $\{\alpha_j\}_{j=1}^{m+1}$ are the cluster fixed effects. Under the assumption that $\mathbb{E}[U_{jt}|D_j, \text{Post}_t] = 0$, Δ can be estimated by

$$\hat{\Delta} = \hat{\theta}_{m+1} - \frac{1}{m} \sum_{j=1}^m \hat{\theta}_j,$$

where $\{\hat{\theta}_j\}_{j=1}^{m+1}$ are the differences in the outcomes before and after treatment. They can be obtained as the estimates of $\{\theta_j\}_{j=1}^{m+1}$ from the following cluster-level regressions

$$Y_{jt} = \alpha_j + \theta_j \text{Post}_t + \epsilon_{jt} \quad (2)$$

for each $j \in \mathcal{J}$. △

Example 2.3 (Two-way fixed effects). Consider the following two-way fixed effects model:

$$Y_{jt} = \alpha_j + \beta_t + \Delta D_j \text{Post}_t + U_{jt}, \quad (3)$$

where $T \geq 2$, $\{\alpha_j\}_{j=1}^{m+1}$ are the cluster fixed effects, $\{\beta_t\}_{t=1}^T$ are the time fixed effects, and $\text{Post}_t \equiv \mathbb{1}[t \geq t_0]$ for some $1 < t_0 \leq T$. Under the assumption that $\mathbb{E}[U_{jt}|D_j, \text{Post}_t] = 0$,

Δ can be estimated by

$$\hat{\Delta} = \hat{\theta}_{m+1} - \frac{1}{m} \sum_{j=1}^m \hat{\theta}_j,$$

using the same cluster-level regression as in (2). \triangle

For more discussion about recent advances in difference-in-differences and two-way fixed effects, see, for instance, the surveys by [de Chaisemartin and D'Haultfœuille \(2022\)](#), [Roth et al. \(2023\)](#), and [Baker et al. \(2025\)](#).

Example 2.4 (Triple differences). Let $C_{ij}, D_{ij} \in \{0, 1\}$ be binary indicators that depend on the individual $i \in \mathcal{I}$ and cluster $j \in \mathcal{J}$. Set $D_{ij} = \mathbb{1}[j = m + 1]$ and define $\text{Post}_t \equiv \mathbb{1}[t = 2]$ as in Example 2.2 with two periods. Assume for each $j \in \mathcal{J}$, there exist units with $C_{ij} = 1$ and units with $C_{ij} = 0$. Consider the following triple differences/difference-in-difference-in-differences model:

$$\begin{aligned} Y_{ijt} = & \beta_0 + \beta_1 C_{ij} + \beta_2 D_{ij} + \beta_3 \text{Post}_t + \beta_4 C_{ij} D_{ij} \\ & + \beta_5 C_{ij} \text{Post}_t + \beta_6 D_{ij} \text{Post}_t + \Delta C_{ij} D_{ij} \text{Post}_t + U_{ijt}. \end{aligned} \quad (4)$$

Under the assumption that $\mathbb{E}[U_{ijt} | C_{ij}, D_{ij}, \text{Post}_t] = 0$, Δ can be estimated by

$$\hat{\Delta} = \hat{\theta}_{m+1} - \frac{1}{m} \sum_{j=1}^m \hat{\theta}_j,$$

where $\{\hat{\theta}_j\}_{j=1}^{m+1}$ are the estimates of $\{\theta_j\}_{j=1}^{m+1}$ from the following cluster-level regressions

$$Y_{ijt} = \alpha_j + \gamma_{C,j} C_{ij} + \gamma_{\text{Post},j} \text{Post}_t + \theta_j C_{ij} \text{Post}_t + \epsilon_{ijt},$$

for each $j \in \mathcal{J}$. See [Olden and Møen \(2022\)](#) for a recent survey on triple difference estimators. \triangle

3 Inference procedure

In this section, we present the main assumptions and our algorithm of conducting inference using the t -test with a single treated cluster. Readers who are interested in applying our test can directly apply Algorithm 3.1. We also present some brief intuition for the validity of our test in Section 3.4 to facilitate the theoretical discussion in Section 4.

3.1 Assumptions

We follow the notation used in Section 2. Let $\{\hat{\theta}_j\}_{j=1}^{m+1}$ be the cluster-level estimators, $\mathcal{J}_0 = \{1, \dots, m\}$ index the control clusters and $\mathcal{J}_1 = \{m+1\}$ index the treated cluster. For simplicity, we make the dependence of $\{\hat{\theta}_j\}_{j=1}^{m+1}$ on the cluster size implicit.

Our procedure requires two main assumptions. First, we assume that an appropriate central limit theorem applies to $\{\hat{\theta}_j\}_{j=1}^{m+1}$ as the sample size within each cluster, denoted by n , goes to infinity. A similar assumption is also imposed in related papers on a fixed number of clusters, such as [Ibragimov and Müller \(2010, 2016\)](#), [Canay et al. \(2017\)](#), [Hagemann \(2022, 2024\)](#) and [Lau \(2025\)](#). Intuitively, this holds when each cluster has a large number of units or consists of a panel with a long time periods. This high-level assumption is formalized as follows.

Assumption 3.1. *The following holds as the sample size within each cluster $n \rightarrow \infty$:*

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_1 - \mu_0 \\ \vdots \\ \hat{\theta}_m - \mu_0 \\ \hat{\theta}_{m+1} - \mu_1 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma), \quad (5)$$

where $\Sigma \equiv \text{diag}(\sigma_1^2, \dots, \sigma_m^2, \sigma_{m+1}^2)$ is an diagonal matrix.

Below we give a few remarks regarding Assumption 3.1. First, it is not necessary for all clusters to have the same size. In cases where clusters have varying sizes, we can define n as the size of the smallest cluster, and our inference essentially requires the sample sizes in all clusters to be large. Second, we assume that the cluster-level estimators are independent, at least asymptotically, so that the asymptotic covariance matrix Σ in (5) is diagonal. This is trivially satisfied if the units are independent across clusters. We emphasize that, while we require independence across clusters, we allow for unknown dependence structure within clusters. Third, we assume that the estimators from the control clusters are consistent for a common parameter μ_0 , which can often be justified by model assumptions or study designs as illustrated in Section 2.

Second, we impose the following relative heterogeneity assumption, which generalizes the relative heterogeneity assumption in [Hagemann \(2024\)](#). Let $\sigma_{(1)} \leq \sigma_{(2)} \leq \dots \leq \sigma_{(m)}$ be the ordered values of $\{\sigma_j\}_{j=1}^m$ in Assumption 3.1.

Assumption 3.2. *For a given $\rho \geq 0$ and a given $k \in \{1, \dots, m\}$, $\sigma_{m+1} \leq \rho \sigma_{(k)}$.*

The above assumption does not require the variances to be known. It only restricts the

relative heterogeneity of the standard deviations between the treated and control clusters. In particular, for any $\rho \geq 0$ and $k \in \{1, \dots, m\}$, it requires the standard deviation of the treated cluster to be less than or equal to ρ times the standard deviations of at least $(m - k + 1)$ control clusters. For example, $k = 1$ requires that the standard deviation of the treated cluster is smaller than or equal to ρ times the standard deviations of all control clusters; when $k = m$, this means that the standard deviation of the treated cluster is smaller than or equal to ρ times the largest standard deviation from the control clusters.

Assumption 3.2 is a relaxed version of Hagemann (2024)'s maximum relative heterogeneity assumption. His assumption is equivalent to Assumption 3.2 with $k = 1$ or $k = 2$. Our test allows a general choice of $k \in \{1, \dots, m\}$. More importantly, as demonstrated in Section 5, the inference can be simultaneously valid for all $k \in \{1, \dots, m\}$. In other words, with additional choices of k , our test can provide more evidences against the null hypothesis than Hagemann (2024)'s test, which focuses on $k = 1$ or 2.

Before introducing our algorithm, we end this subsection with some discussion on choosing (ρ, k) in Assumption 3.2. If the researcher believes the standard deviation of the treated cluster cannot be larger than the standard deviations of the control clusters, then the researcher can set $k = 1$ and $\rho = 1$.

On the other hand, if the researcher does not want to commit to a particular value of (ρ, k) , the researcher can perform simultaneous inference and report the largest value of ρ such that the null can be rejected for each k . This approach can also be interpreted as finding the largest ρ such that the conclusion changes, which is related to the idea of finding breakdown points/frontiers in econometrics (Horowitz and Manski, 1995; Kline and Santos, 2013; Masten and Poirier, 2020). We demonstrate this through two applications in Section 7.

3.2 Algorithm

The goal is to test the following hypothesis:

$$H_0: \mu_1 = \mu_0 \quad \text{vs} \quad H_1: \mu_1 \neq \mu_0. \quad (6)$$

Our procedure of conducting inference with a single treated cluster is as follows.

Algorithm 3.1.

Inputs:

- Significance level $\alpha \in (0, \frac{1}{2})$.

- The parameters on relative heterogeneity from Assumption 3.2, i.e., (ρ, k) .
- The cluster-level estimators $\{\hat{\theta}_j\}_{j=1}^{m+1}$.

Steps:

Step 1 (Test statistic): Compute the t -statistic:

$$\hat{T}_m \equiv \frac{\hat{\theta}_{m+1} - \bar{\hat{\theta}}_m}{\hat{S}_m}, \quad (7)$$

where $\bar{\hat{\theta}}_m \equiv \frac{1}{m} \sum_{j=1}^m \hat{\theta}_j$ and $\hat{S}_m^2 \equiv \frac{1}{m-1} \sum_{j=1}^m (\hat{\theta}_j - \bar{\hat{\theta}}_m)^2$ represent the sample average and sample variance of the estimators from the control clusters, respectively.

Step 2 (Critical value): We have tabulated many empirically-relevant critical values in Table 1. These critical values can be immediately used. More generally, the critical value $\text{cv}_{m,\alpha,k,\rho}$ can be computed as follows.

- **Case 1:** If $k = 1$ and α is below the thresholds in Table 2, use the closed-form expression in (17). Compute the closed-form critical value as

$$\text{cv}_{m,\alpha,k,\rho} \equiv \sqrt{\rho^2 + \frac{1}{m}} t_{m-1, \frac{\alpha}{2}},$$

where $t_{m-1, \frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ -th quantile of the t -distribution with $(m - 1)$ degrees of freedom.

- **Case 2:** Search $\text{cv}_{m,\alpha,k,\rho}$ such that $p_m(\text{cv}_{m,\alpha,k,\rho}; k, \rho)$ defined in Theorem 4.3 ahead is at most α . This requires solving one-dimensional optimization problems.

Step 3 (Decision): Reject (6) if $|\hat{T}_m| > \text{cv}_{m,\alpha,k,\rho}$. ■

3.3 Computing p -value and confidence interval

To compute the p -value, the researcher does not need to run the entire Algorithm 3.1. In particular, the researcher only needs to compute the test statistic \hat{T}_m in Step 1 of Algorithm 3.1 and substitute it into the p_m function in Step 2 of Algorithm 3.1 and return the p -value as $p_m(\hat{T}_m; k, \rho)$.

The researcher can take the critical value $\text{cv}_{m,\alpha,k,\rho}$ from Step 2 of Algorithm 3.1 and return $(\hat{\theta}_{m+1} - \bar{\hat{\theta}}_m) \pm \text{cv}_{m,\alpha,k,\rho} \hat{S}_m$ as an $(1 - \alpha)$ confidence interval for $\mu_1 - \mu_0$.

3.4 Discussion

Algorithm 3.1 is a computationally simple three-step procedure. The first step is to compute the usual t -statistic using the cluster-level estimators.

The second step finds the critical value that depends on (m, α, k, ρ) . This step aims to find the critical value $cv_{m, \alpha, k, \rho}$ such that the procedure controls size for any configurations of $\{\sigma_j\}_{j=1}^{m+1}$ that satisfy Assumption 3.2. In the first case, a closed-form critical value is available when α is below a certain threshold in Table 2. If α does not satisfy the condition, we can compute the critical value $cv_{m, \alpha, k, \rho}$ in the second step through one-dimensional optimization.

The reason for having a closed-form expression for the critical value in case 1 of Step 2 of Algorithm 3.1 is as follows. For $k = 1$ and any given (m, ρ) , for a sufficiently small significance level α (including many conventional choices), we find that the maximum rejection probability over all configurations of $\{\sigma_j\}_{j=1}^{m+1}$ is achieved when $\sigma_{m+1} = \rho\sigma_j$ for all $j = 1, \dots, m$. The proof is nontrivial and we explain the technical details in Sections 4. Knowing when the maximum rejection probability is achieved, we are able to derive a closed-form expression for the critical value using properties of the t -distribution.

Case 2 of Step 2 is in fact a general version that holds for any (m, α, k, ρ) . For a general value of $k \in \{1, \dots, m\}$, although we cannot find the exact configuration of $\{\sigma_j\}_{j=1}^{m+1}$ that achieves the desired level of rejection probability, we are able to substantially reduce the set of possible values. In particular, there are at most m^2 possible cases for the worst-case configuration of $\{\sigma_j\}_{j=1}^{m+1}$, each of which involves at most one unknown parameter. Therefore, we can find the maximum rejection probability through one-dimensional optimization. In addition, we have a good initial guess for the critical value. To facilitate the use of the test in practice, we have tabulated many critical values for researchers' use in Table 1.

The last step rejects if the absolute value of the test statistic is above the critical value.

4 Theory

In this section, we develop the theory that justifies the validity of Algorithm 3.1. In Section 4.1, we show that the large-sample behavior of the test can be studied via a fixed number of normal random variables. In Section 4.2, we show the general theory on computing the maximum rejection probability of the test statistic. This result is used to search for a critical value such that the test is valid. We show in Section 4.3 that we can get a closed-

form expression for the critical value as in Step 2 of Algorithm 3.1 when $k = 1$ and the significance level is not “too large.” In Section 4.4, we discuss the power of our t -test.

4.1 General framework with normal means

In this subsection, we show that under Assumption 3.1, studying the behavior of the t -statistic in equation (7) of Algorithm 3.1 when n is large is equivalent to studying the t -statistic of suitably defined $(m + 1)$ normal random variables. Formally, consider the following assumption on the normal random variables.

Assumption 4.1. Let $\{\psi_j\}_{j=1}^{m+1}$ be $m + 1$ independent random variables, where $\psi_j \sim \mathcal{N}(0, \sigma_j^2)$ for $1 \leq j \leq m$ and $\psi_{m+1} \sim \mathcal{N}(\delta, \sigma_{m+1}^2)$.

Compared to the notation in Section 3.1, $\{\psi_j\}_{j=1}^m$ above correspond to the m control clusters and ψ_{m+1} corresponds to the single treated cluster. As before, the above assumption does not require us to know the variances $\{\sigma_j\}_{j=1}^{m+1}$. The variances can be arbitrarily heterogeneous as long as they satisfy the relative heterogeneity assumption stated in Assumption 3.2. Analogous to (7), we define the following t -statistic based on these normal random variables:

$$T_m \equiv \frac{\psi_{m+1} - \bar{\psi}_m}{S_m}, \quad (8)$$

where $\bar{\psi}_m \equiv \frac{1}{m} \sum_{j=1}^m \psi_j$ and $S_m^2 \equiv \frac{1}{m-1} \sum_{j=1}^m (\psi_j - \bar{\psi}_m)^2$ denote the sample mean and variance for the control clusters, respectively.

The theorem below shows that, to achieve the desired size asymptotically, it suffices to consider a stylized setting with normally distributed random variables. Note that \hat{T}_m in (7) depends on the sample size within each cluster n implicitly. In addition, we allow μ_1 and μ_0 to vary with the sample size n as well, which can facilitate power investigation under local alternatives.

Theorem 4.1. Let $m \geq 2$. Suppose that Assumption 3.1 holds, $\sqrt{n}(\mu_1 - \mu_0) \rightarrow \delta$ as $n \rightarrow \infty$ for some $\delta \in \mathbb{R}$, and the variances $\{\sigma_j^2\}_{j=1}^{m+1}$ are not all zero. Then, for any $c > 0$ and $c \neq m^{-1/2}$,²

$$\lim_{n \rightarrow \infty} \mathbb{P}[|\hat{T}_m| > c] = \mathbb{P}[|T_m| > c],$$

where T_m is the t -statistic defined in (9).

²We impose $c \neq m^{-1/2}$ to avoid the case where $|T_m|$ has a positive point mass at c . This technical requirement excluding a single value of c generally does not affect the practical use of our test, since the desired critical value at a usual significance level is greater than $m^{-1/2}$. Specifically, as discussed later in Remark 4.4, the rejection probability $\mathbb{P}[|T_m| > c]$ for any $c < m^{-1/2}$ can be as large as 1, and the rejection probability $\mathbb{P}[|T_m| > c]$ for c close to but larger than $m^{-1/2}$ can at least be about 0.5.

Note that in the asymptotics of Theorem 4.1, the limit is taken as n goes to infinity, while the number of clusters $(m + 1)$ is fixed. In addition, we allow a general δ for the difference between μ_1 and μ_0 to facilitate the later discussion on the power of the test. To derive valid tests, it suffices to consider the case where $\mu_1 = \mu_0$ and $\delta = 0$. Specifically, to derive large-sample valid t -test with a single treated cluster and a finite number of control clusters, it suffices to test the following null hypothesis in the stylized setting with exactly normal observations:

$$\bar{H}_0 : \delta = 0 \quad \text{vs} \quad \bar{H}_1 : \delta \neq 0. \quad (9)$$

In the remainder of this section, we will study valid t -test for (9) in this stylized setting.

Remark 4.1. If we are interested in an one-sided alternative, such as $\mu_1 > \mu_0$, we can reject the null hypothesis if $\hat{T}_m > c$ for some $c > 0$. Moreover, under the same condition as in Theorem 4.1, $\lim_{n \rightarrow \infty} \mathbb{P}[\hat{T}_m > c] = \mathbb{P}[T_m > c] = \frac{1}{2} \mathbb{P}[|T_m| > c]$. Therefore, the critical value for a level- α one-sided test can be derived from the critical value for a level- (2α) two-sided test, for any $\alpha \in [0, \frac{1}{2}]$. ■

Remark 4.2. The setup in this paper also works when there is a single control cluster and at least two treated clusters. All the results would follow by labeling the first m clusters as the treated clusters and the $(m + 1)$ th cluster as the control cluster. ■

Remark 4.3. Although we focus on a fixed m , we discuss in Appendix A.3 on the behavior of the test when m is large, where we derive a closed-form valid test and approximate its power. ■

4.2 Valid t -test

The key to constructing a valid t -test is to find an appropriate critical value $\text{cv}_{m,\alpha,k,\rho}$ for Step 2 of Algorithm 3.1. This critical value $\text{cv}_{m,\alpha,k,\rho}$ has to be chosen such that for any $\{\sigma_j\}_{j=1}^{m+1}$ that satisfy Assumption 3.2, $\mathbb{P}[|T_m| > \text{cv}_{m,\alpha,k,\rho}]$ under the null hypothesis (9) is less than or equal to a given significance level α . In the following, we will consider the maximum rejection probability at any critical value c .

Note that when the treated standard deviation is much larger than the control standard deviations, i.e., $\sigma_{m+1} \gg \max_{1 \leq j \leq m} \sigma_j$, the rejection probability $\mathbb{P}[|T_m| > c]$ will be close to 1 for any $c > 0$, under which we cannot derive a meaningful t -test. Therefore, Assumption 3.2 on the relative heterogeneity of $\{\sigma_j\}_{j=1}^{m+1}$ is in some sense necessary.

For descriptive convenience, we introduce $\mathcal{S}_m(k, \rho)$ to denote all possible standard de-

viations that satisfy Assumption 3.2 for a given $m, \rho \geq 0$ and $k \in \{1, \dots, m\}$ ³:

$$\mathcal{S}_m(k, \rho) \equiv \{(\sigma_1, \dots, \sigma_m, \sigma_{m+1}) \in \mathbb{R}_{\geq 0}^{m+1} : \sigma_{m+1} \leq \rho \sigma_{(k)}\}. \quad (10)$$

Using the above notation, our goal of finding the maximum rejection probability as described above can be formalized as follows

$$p_m(c; k, \rho) \equiv \sup_{(\sigma_1, \dots, \sigma_m, \sigma_{m+1}) \in \mathcal{S}_m(k, \rho)} \mathbb{P}_0[|T_m| > c], \quad (11)$$

where the \mathbb{P}_0 notation with subscript 0 indicates that the null hypothesis (9) holds and $\delta = 0$. To conduct a valid test at any given significance level $\alpha \in (0, 1)$, it suffices to find the critical value c such that $p_m(c; k, \rho) \leq \alpha$ and ideally $p_m(c; k, \rho) = \alpha$. On the other hand, if the goal is to compute a p -value, then one can directly compute $p_m(|T_m|; k, \rho)$ without searching for the required critical value as discussed in Section 3.3. That is, when we set c to be the observed absolute value of the t -statistic, $p_m(c; k, \rho)$ gives a valid p -value for testing the null hypothesis in (9).

In the following, we consider two cases depending on the standard deviation for the treated cluster σ_{m+1} . Section 4.2.1 considers $\sigma_{m+1} = 0$ and shows that the corresponding maximum rejection probability has a closed-form solution that can be easily computed. Section 4.2.2 considers $\sigma_{m+1} > 0$ and shows that we have an integral representation for the rejection probability. This facilitates its numerical calculation at any given values of $\{\sigma_j\}_{j=1}^{m+1}$. However, directly evaluating the optimization problem in (11) generally results in solving an m -dimensional optimization problem. We discuss how this can be reduced to solving multiple one-dimensional optimization problems in Section 4.2.3.

4.2.1 Case 1: $\sigma_{m+1} = 0$

When $\sigma_{m+1} = 0$, regardless of the choices of the relative heterogeneity parameters ρ and k , the control standard deviations $\{\sigma_j\}_{j=1}^m$ can take arbitrary values in $\mathbb{R}_{\geq 0}^m$. Moreover, in this case, except for a multiplicative constant scaling factor of \sqrt{m} in the t -statistic, our t -statistic essentially reduces to a one-sample t -statistic, and our t -test is equivalent to testing whether the mean of the control clusters is equal to zero. From Bakirov and Székely (2006), we know that the maximum rejection probability with a zero σ_{m+1} has the

³With a slight abuse of notation, we also use $\{\sigma_j\}_{j=1}^{m+1}$ to denote generic values of the standard deviations. Note that $\{\sigma_j\}_{j=1}^{m+1}$ in Assumptions 3.1 and 3.2 represent the true (asymptotic) standard deviations for all the clusters.

following form:

$$p_{m,0}(c) = \sup_{\substack{(\sigma_1, \dots, \sigma_m) \in \mathbb{R}_{\geq 0}^m \\ \text{and } \sigma_{m+1}=0}} \mathbb{P}_0[|T_m| > c] = \max_{R_m(c) < j \leq m} \mathbb{P} \left[|t_{j-1}| > \sqrt{\frac{(j-1)R_m(c)}{j - R_m(c)}} \right], \quad (12)$$

where $R_m(c) \equiv \frac{m^2 c^2}{m c^2 + m - 1}$, and t_{j-1} denotes a random variable following the t -distribution with degrees of freedom $(j-1)$. In (12), the maximum rejection probability is obtained when the control standard deviations $\{\sigma_j\}_{j=1}^m$ are either zero or take some common positive value. It can be efficiently computed by calculating the tail probabilities of at most $(m-1)$ t -distributions with various degrees of freedom.

It follows that when $\rho = 0$, we must have $\sigma_{m+1} = 0$, and the maximum rejection probability of our t -test becomes the same as (12), i.e., $p_m(c; k, 0) = p_{m,0}(c)$ for any $c > 0$ and $k \in \{1, \dots, m\}$. Thus, in the remainder of this section, we focus on $\rho > 0$.

Remark 4.4. From (12) and as commented in Bakirov and Székely (2006), $p_{m,0}(c) = 1$ for $0 < c < m^{-1/2}$ and $p_{m,0}(c) = 0.5$ for $c = m^{-1/2}$. Furthermore, by the right-continuous property of distribution functions, we can verify that $p_{m,0}(c) \rightarrow 0.5$ as c approaches $m^{-1/2}$ from the right. These observations mean that, regardless of the values of ρ and k , the maximum rejection probability of our t -test in (11) is 1 when $0 < c < m^{-1/2}$, at least 0.5 when $c = m^{-1/2}$, and at least about 0.5 when c is greater than but close to $m^{-1/2}$. Thus, it is generally innocuous to assume $c \neq m^{-1/2}$ (or even $c > m^{-1/2}$) as in Theorems 4.1 and later in 4.3 for most conventional significance levels. ■

4.2.2 Case 2: $\sigma_{m+1} > 0$

We now consider the case where $\sigma_{m+1} > 0$. As shown in the following lemma, the rejection probability $\mathbb{P}_0[|T_m| > c]$ can be written as an integral. This not only facilitates numerical computation, but is also crucial for our later theoretical investigation.

Lemma 4.1. Suppose that $\sigma_{m+1} > 0$, and define $\gamma_i \equiv \frac{\sigma_i}{\sigma_{m+1}}$ for $i = 1, 2, \dots, m$. The rejection probability $\mathbb{P}_0[|T_m| > c]$ can be written as:

$$\mathbb{P}_0[|T_m| > c] = \bar{p}_m(c; \gamma_1, \dots, \gamma_m) \equiv \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{s^{\frac{m-1}{2}}}{[-g_c(-s)]^{\frac{1}{2}}} ds. \quad (13)$$

where $\theta_{m+1} \in [-m - \max_{1 \leq i \leq m} \gamma_i^2, -m]$ is the unique negative root of $g_c(\theta)$ defined below,

$$g_c(\theta) \equiv -(m + \theta) \prod_{i=1}^m (\kappa \gamma_i^2 - \theta) + \left(\kappa + \frac{\kappa + 1}{m} \theta \right) \cdot \sum_{i=1}^m \left[\gamma_i^2 \prod_{j \neq i} (\kappa \gamma_j^2 - \theta) \right], \quad (14)$$

and $\kappa \equiv \frac{mc^2}{m-1}$.

From Lemma 4.1, when $\sigma_{m+1} > 0$, the rejection probability in (13) depends only on the ratios between the control standard deviations and the treated standard deviation. The relative heterogeneity assumption in Assumption 3.2 equivalently assumes that at least $(m - k + 1)$ of $\{\gamma_j\}_{j=1}^m$ is greater than or equal to ρ^{-1} . To find the maximum rejection probability, we need to solve an m -dimensional optimization over $(\gamma_1, \dots, \gamma_m) \in [0, \infty)^{k-1} \times [\rho^{-1}, \infty)^{m-k+1}$.⁴ Optimizing this integral directly can be computationally challenging even for a moderate m . As demonstrated in the next subsection, for any $m \geq 2$, we can substantially simplify the m -dimensional optimization problem into multiple one-dimensional optimization problems. The reformulated problem can often be efficiently solved numerically.

4.2.3 Reformulating the problem of computing the maximum rejection probability

In this subsection, we discuss how to reduce the optimization for the maximum rejection probability in (11) into multiple one-dimensional optimization problems. We summarize the key ideas here and defer the technical lemmas to Appendix A.1 of this main text.

First, Lemma A.1 shows that the maximum rejection probability in (11) must be achieved at some finite values of $(\sigma_1, \dots, \sigma_m, \sigma_{m+1}) \in \mathcal{S}_m(k, \rho)$. This means we do not need to worry about the boundary case where some of $\{\sigma_j\}_{j=1}^{m+1}$ approach infinity. Next, Lemma A.2 shows the necessary conditions for any $(\sigma_1, \dots, \sigma_m, \sigma_{m+1}) \in \mathcal{S}_m(k, \rho)$ to be a maximizer. We summarize the implications of Lemma A.2 in the following theorem.

Theorem 4.2. *For any given $m \geq 2$, $k \in \{1, \dots, m\}$, $\rho \geq 0$, $c > 0$ and $c \neq m^{-1/2}$, the maximum rejection probability in (11) under Assumption 3.2 must be attained at some $\{\sigma_j\}_{j=1}^{m+1} \in \mathcal{S}_m(k, \rho)$ that satisfy one of the following forms:⁵*

- (i) $\sigma_{m+1} = 0$, and for $1 \leq j \leq m$, σ_j is either 0 or some common value;
- (ii) $\sigma_{m+1} = \rho$, and for $1 \leq j \leq m$, σ_j is either 0, or 1, or some common value.

⁴Note that the rejection probability is invariant to any permutations of the $\{\gamma_j\}_{j=1}^m$. We can, for example, assume that the last $(m - k + 1)$ of them is no less than ρ^{-1} without loss of generality.

⁵The result in (i) follows from Bakirov and Székely (2006) and implies the maximum rejection probability when $\sigma_{m+1} = 0$ as shown in (12).

Note that 0 is always a boundary value of σ_j for $1 \leq j \leq m$, and 1 is also a boundary value for some σ_j when $\sigma_{m+1} = \rho$ and Assumption 3.2 holds. Therefore, Theorem 4.2 essentially indicates that, at the maximizer of the rejection probability in (11), each σ_j must either lie on the boundary or share a common value for all $1 \leq j \leq m$. Importantly, Theorem 4.2 explains why we can simplify the optimization for the rejection probability into multiple one-dimensional optimization problems. This is because for all possible maximizers shown in Theorem 4.2, there is only one unknown value, which is the common value of the $\{\sigma_j\}_{j=1}^m$ that are not on the boundary. Note that the maximum rejection probability in case (i) with $\sigma_{m+1} = 0$ can be efficiently computed as shown in (12). In the following, we will therefore focus on the optimization for the rejection probability in case (ii) with $\sigma_{m+1} > 0$.

Specifically, under Assumption 3.2 with any $k \in \{1, \dots, m\}$ and $\rho > 0$, if $(\sigma_1, \dots, \sigma_m, \sigma_{m+1}) \in \mathcal{S}_m(k, \rho)$ is the maximizer for the rejection probability and $\sigma_{m+1} > 0$, then, for all $1 \leq j \leq m$, $\gamma_j \equiv \frac{\sigma_j}{\sigma_{m+1}}$ is either on the boundary (equals 0 or ρ^{-1}), or takes some common value $\gamma \geq 0$. Motivated by this, with a slight abuse of notation, we introduce $\bar{p}_m(c; \rho, \gamma; m_1, m_0)$ to denote the value of $\bar{p}_m(c; \gamma_1, \dots, \gamma_m)$ in (13) when m_1 of $\{\gamma_j\}_{j=1}^m$ equal ρ^{-1} , m_0 of them equal zero, and the remaining equal a common value γ . That is,

$$\bar{p}_m(c; \rho, \gamma; m_1, m_0) \equiv \bar{p}_m(c; \rho^{-1} \mathbf{1}_{m_1}^\top, \mathbf{0}_{m_0}^\top, \gamma \mathbf{1}_{m-m_1-m_0}^\top), \quad (15)$$

where $0 \leq m_0, m_1 \leq m$ and $m_0 + m_1 \leq m$.

Next, define the following as the supremum of $\bar{p}_m(c; \rho, \gamma; m_1, m_0)$ over γ , with the support of γ depending on m_1 and k :

$$\tilde{p}_m(c; k, \rho; m_1, m_0) \equiv \sup_{\gamma \in [\underline{\rho}, \infty)} \bar{p}_m(c; \rho, \gamma; m_1, m_0), \quad (16)$$

where $\underline{\rho} = 0$ if $m_1 \geq m - k + 1$ and $\underline{\rho} = \rho^{-1}$ if $m_1 < m - k + 1$.

With the above notations, we now state our main theorem of how to evaluate the maximum rejection probability for a given critical value c , heterogeneity parameters (k, ρ) and the number of clusters m .

Theorem 4.3. *For any given $m \geq 2$, $k \in \{1, \dots, m\}$, $c > 0$ and $c \neq m^{-1/2}$, the maximum*

rejection probability in (11) under Assumption 3.2 can be written as⁶

$$p_m(c; k, \rho) \equiv \begin{cases} \max \left\{ p_{m,0}(c), \max_{0 \leq m_0 \leq k-1, 0 \leq m_1 \leq m-m_0} \tilde{p}_m(c; k, \rho; m_1, m_0) \right\} & \text{if } \rho > 0 \\ p_{m,0}(c) & \text{if } \rho = 0 \end{cases},$$

where $p_{m,0}(c)$ is defined in (12) and $\tilde{p}_m(c; k, \rho; m_1, m_0)$ is defined in (16).

From Theorem 4.3, the key to obtaining the maximum rejection probability is to solve the optimization problem in (16) for all combinations of (m_1, m_0) . Importantly, for each given (m_1, m_0) , the optimization in (16) is an one-dimensional optimization problem, which is computationally much simpler than the original optimization in (11), and we can solve it using numerical optimization. In particular, for any given $k \in \{1, \dots, m\}$, we need to solve at most $\frac{1}{2}k(2m+1-k)$ one-dimensional optimization problems⁷, which is m when $k = 1$, $2m-1$ when $k = 2$, \dots , and $\frac{1}{2}m(m+1)$ when $k = m$. Moreover, at a given $m, \rho \geq 0$ and $c > 0$, to find the maximum rejection probability over all $k \in \{1, \dots, m\}$, we need to solve at most $m(m+1)$ one-dimensional optimization problems of the form (16), because the optimizations required for different values of k overlap with each other.

We illustrate the above theorem through the following examples. Example 4.1 is about $k = 1$ and Example 4.2 is about $k = 2$.

Example 4.1. Let $m = 6$, $\rho = 2$, $\alpha = 0.05$, $k = 1$, and $c = t_{m-1, 1-\frac{\alpha}{2}} \sqrt{\rho^2 + \frac{1}{m}}$ where $t_{m-1, 1-\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ -th quantile of the t -distribution with $(m-1)$ degrees of freedom. We show the various functions from Theorem 4.3 in Figure 1.

A few observations from Figure 1 are as follows. First, the curves show $\bar{p}_m(c; \rho, \gamma; m_1, m_0)$ as defined in (15) against γ , for $m_1 = 0, \dots, 5$. Recall this function means that among $\{\gamma_j\}_{j=1}^6$, m_1 of them equals ρ^{-1} and the remaining of them equals γ . We set $\gamma = e^x$ in order to display the behavior of large γ . Each colored curve corresponds to a specific value of m_1 . It can be seen that $\bar{p}_m(c; \rho, \gamma; m_1, m_0)$ decreases as γ increases for each m_1 .

The black dotted horizontal line plots $p_{m,0}(c)$ defined in (12). It shows the maximum rejection probability when $\sigma_{m+1} = 0$. Here, it is much smaller than 0.05.

As $\gamma \rightarrow \infty$, $\bar{p}_m(c; \rho, \gamma; m_1, m_0)$ converges to values less than or equal to $p_{m,0}(c)$. This is not surprising, as in such scenarios, σ_{m+1} is much smaller than some control standard deviations, essentially approximating the case where $\sigma_{m+1} = 0$. \triangle

⁶From Remark 4.4, $p_m(c; k, \rho) = 1$ when $0 < c < m^{-1/2}$. For convenience, we also define $p_m(c; k, \rho) = 1$ at $c = 0$ and $c = m^{-1/2}$, so that $p_m(c; k, \rho)$ upper bounds the rejection probability for all $c \geq 0$.

⁷When $m_1 + m_0 = m$, (15) no longer depends on γ , and thus no optimization over γ is needed.

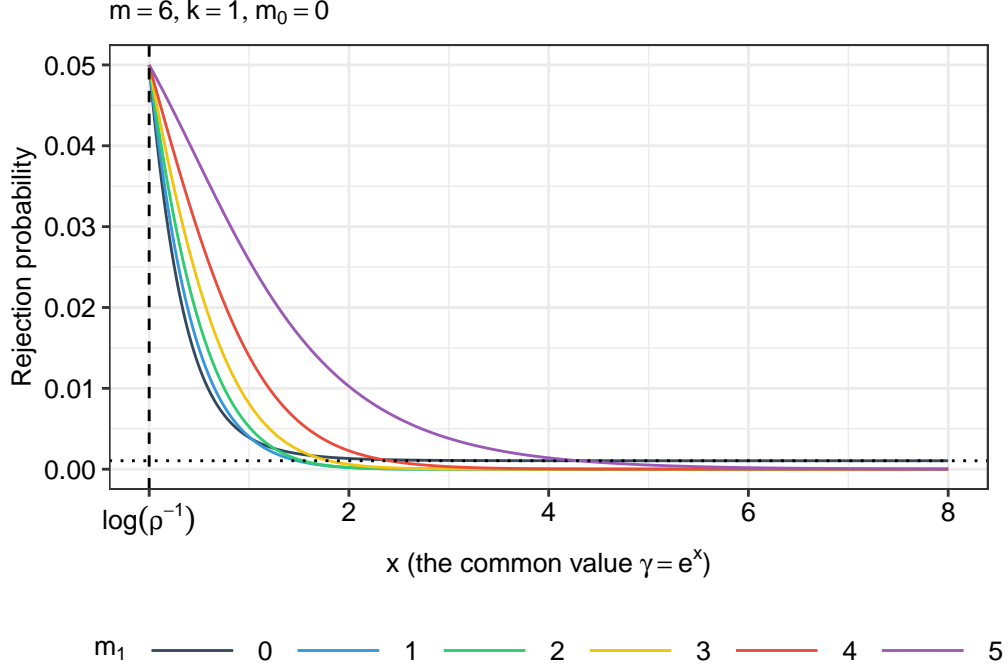


Figure 1: This above shows $\bar{p}_m(c; \rho, \gamma; m_1, m_0)$ against γ for various values of m_1 with $m = 6$, $k = 1$ and $m_0 = 0$. The vertical dashed line represents $\gamma = \rho^{-1}$. The horizontal dotted line represents $p_{m,0}(c)$. See Theorem 4.3 for the definitions of the functions.

Example 4.2. Consider the same setup as in Example 4.1 but with $k = 2$ and set the value of c such that the maximum rejection probability equals $\alpha = 0.05$. Figure 2 shows the rejection probability under different combinations of γ , m_0 and m_1 , where m_0 here can only take the values 0 or 1.

As can be seen from the figure, the maximum rejection probability is achieved when one of the $\{\gamma_j\}_{j=1}^m$ equals 0 and the remaining $(m - 1)$ terms from $\{\gamma_j\}_{j=1}^m$ equals ρ^{-1} . For instance, in the $m_0 = 0$ panel, the purple line represents $m_1 = 5$ terms of $\{\gamma_j\}_{j=1}^m$ equals ρ^{-1} . For this line, the maximum rejection probability is achieved when the remaining $m - m_1 = 1$ term equals $\gamma = 0$. Similarly, in the $m_0 = 1$ panel, exactly one of $\{\gamma_j\}_{j=1}^m$ is equal to 0. Each of the lines shows that the maximum rejection probability is achieved when all the remaining terms equal ρ^{-1} . \triangle

Remark 4.5. As discussed in Appendix A.3, when $m \rightarrow \infty$, under certain regularity conditions, the maximum rejection probability $p_m(c; k, \rho)$ for any c, k and $\rho > 0$ is achieved when $\sigma_{m+1} = 1$, $(k - 1)$ of $\{\sigma_j\}_{j=1}^m$ are 0, and the remaining $(m - k + 1)$ of $\{\sigma_j\}_{j=1}^m$ are equal to ρ^{-1} . This is indeed the case for both Examples 4.1 and 4.2. Based on

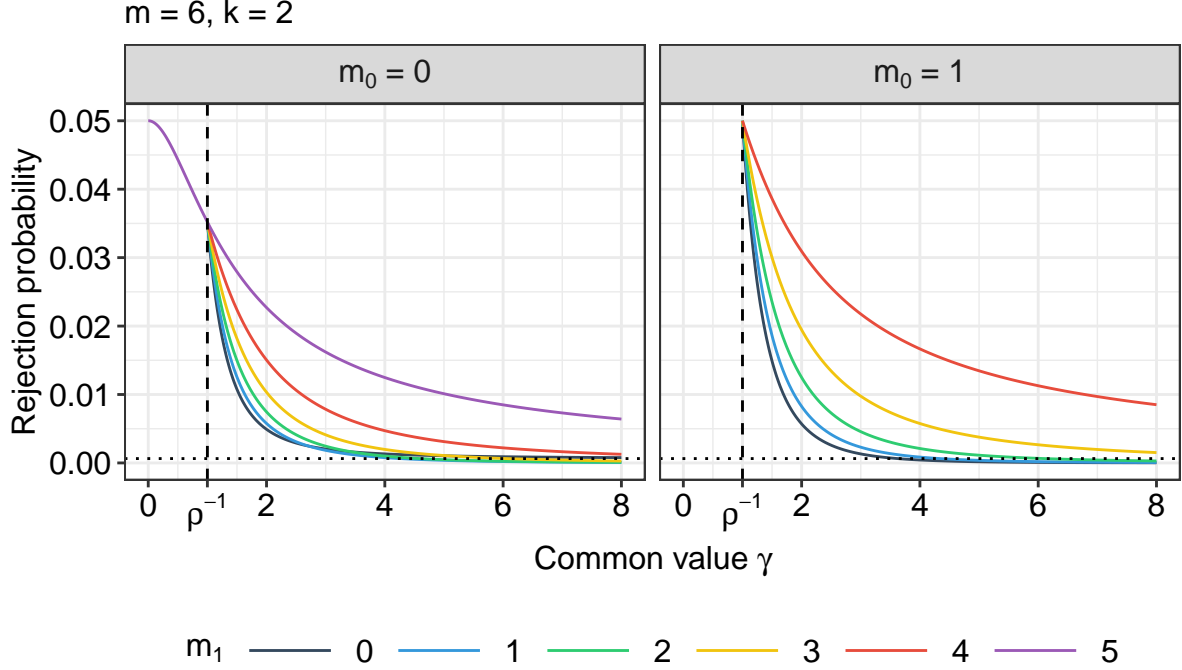


Figure 2: This above shows $\bar{p}_m(c; \rho, \gamma; m_1, m_0)$ against γ for various m_0 and m_1 with $m = 6$ and $k = 1$. See the caption for Figure 1 for more details.

this intuition, we can first use this configuration of $\{\sigma_j\}_{j=1}^{m+1}$ to get a candidate threshold c such that $\mathbb{P}_0[|T_m| > c] = \alpha$, where α is the significance level of interest. We can then verify whether this candidate threshold c achieves the desired type-I error control by solving the optimization in Theorem 4.3. ■

Finally, we report the critical values for different numbers of control clusters m and heterogeneity parameters ρ for $\alpha = 0.01$ and $\alpha = 0.05$ in Table 1 when $k = 1$. We report the critical values for $k = 2$ in the supplementary material.

4.3 Closed-form valid t -test when $k = 1$

In this subsection, we consider the case where $k = 1$ in Assumption 3.2. This means that $\sigma_{m+1} \leq \rho\sigma_j$ for $j = 1, \dots, m$, i.e., the treated standard deviation is smaller than or equal to ρ times each of the control standard deviations.

In this case, we can obtain closed-form solutions for the maximum rejection probability when the threshold c is large enough. Equivalently, this corresponds to testing at a significance level that is not “too large.” The theorem is stated as follows:

Table 1: Critical values for different values of α , m and ρ for $k = 1$.

$\rho \backslash m$	$\alpha = 0.01$						$\alpha = 0.05$					
	5	10	15	20	25	50	5	10	15	20	25	50
0.2	2.256	1.216	0.972	0.858	0.791	0.656	1.360	0.846	0.700	0.628	0.584	0.492
0.4	2.762	1.657	1.417	1.311	1.251	1.137	1.666	1.153	1.021	0.959	0.923	0.853
0.6	3.445	2.204	1.944	1.832	1.769	1.652	2.078	1.534	1.401	1.340	1.305	1.239
0.8	4.220	2.796	2.502	2.376	2.306	2.177	2.545	1.946	1.803	1.739	1.702	1.633
1.0	5.044	3.408	3.074	2.932	2.852	2.707	3.041	2.373	2.215	2.145	2.105	2.030
1.2	5.896	4.033	3.654	3.492	3.403	3.238	3.556	2.807	2.633	2.555	2.511	2.428
1.4	6.767	4.664	4.238	4.056	3.955	3.771	4.081	3.247	3.053	2.967	2.919	2.828
1.6	7.649	5.300	4.825	4.622	4.510	4.305	4.613	3.689	3.476	3.381	3.328	3.228
1.8	8.539	5.939	5.413	5.189	5.065	4.839	5.150	4.134	3.900	3.796	3.738	3.628
2.0	9.436	6.580	6.003	5.758	5.622	5.373	5.690	4.581	4.325	4.212	4.148	4.029
2.2	10.336	7.223	6.594	6.326	6.179	5.908	6.233	5.028	4.751	4.628	4.559	4.430
2.4	11.240	7.867	7.186	6.896	6.736	6.443	6.778	5.476	5.177	5.045	4.971	4.831
2.6	12.146	8.512	7.778	7.466	7.294	6.978	7.325	5.925	5.604	5.462	5.382	5.233
2.8	13.055	9.157	8.371	8.036	7.851	7.513	7.873	6.374	6.031	5.879	5.794	5.634
3.0	13.965	9.804	8.964	8.607	8.409	8.049	8.421	6.824	6.458	6.296	6.205	6.035
3.2	14.876	10.450	9.557	9.177	8.968	8.584	8.971	7.274	6.886	6.714	6.617	6.437
3.4	15.789	11.097	10.150	9.748	9.526	9.120	9.521	7.725	7.313	7.132	7.029	6.838
3.6	16.702	11.744	10.744	10.319	10.085	9.655	10.072	8.175	7.741	7.549	7.441	7.240
3.8	17.616	12.392	11.338	10.890	10.643	10.191	10.623	8.626	8.169	7.967	7.854	7.642
4.0	18.531	13.040	11.932	11.462	11.202	10.727	11.175	9.077	8.597	8.385	8.266	8.043
4.2	19.447	13.688	12.526	12.033	11.760	11.262	11.727	9.528	9.025	8.803	8.678	8.445
4.4	20.362	14.336	13.121	12.604	12.319	11.798	12.279	9.979	9.453	9.221	9.091	8.847
4.6	21.279	14.985	13.715	13.176	12.878	12.334	12.832	10.430	9.882	9.639	9.503	9.248
4.8	22.195	15.633	14.310	13.747	13.437	12.869	13.385	10.882	10.310	10.057	9.915	9.650
5.0	23.112	16.282	14.904	14.319	13.996	13.405	13.938	11.333	10.738	10.476	10.328	10.052

Theorem 4.4. For any given $m \geq 4$, $\rho > 0$, $c > \sqrt{\frac{3(m-1)}{m(m-3)}}$, define:

$$\begin{aligned} \bar{H}_m(c, \rho) \equiv & \max \left\{ \frac{3(m\rho^2 + 1)}{m\rho^2 + \kappa + 1}, \frac{2\kappa + 3}{\kappa + 1} \right\} \\ & + \frac{1 - \tau}{1 - \tau + \min\{(1 - 2\tau)\kappa\underline{Z} - \frac{1}{2}, 0\}} - \frac{m\kappa}{m\rho^2 + \kappa + 1} - 1, \end{aligned}$$

where $\kappa \equiv \frac{mc^2}{m-1}$ and $\tau \equiv \frac{\kappa+1}{m\kappa}$ are determined by (m, c) , and $\underline{Z} \equiv \frac{1}{2 \cdot \max\{m\rho^2+1, \kappa+2\}}$ is determined by (m, c, ρ) . Under the above conditions and notations, the following statements hold.

- (i) $\bar{H}_m(c, \rho)$ is decreasing in c , and $\lim_{c \rightarrow \infty} \bar{H}_m(c, \rho) < 0$.

- (ii) Let $\underline{c}_{m,\rho} \equiv \inf\{c > \sqrt{\frac{3(m-1)}{m(m-3)}} : \bar{H}_m(c, \rho) \leq 0\}$, which must be finite. Then, for any $c \geq \underline{c}_{m,\rho}$, the maximum rejection probability $p_m(c; 1, \rho)$ under Assumption 3.2 with $k = 1$ and the given (ρ, m, c) has the following equivalent form:

$$p_m(c; 1, \rho) = \mathbb{P} \left[|t_{m-1}| \sqrt{\rho^2 + \frac{1}{m}} > c \right].$$

Theorem 4.4 gives a closed-form solution for the maximum rejection probability under the relative heterogeneity assumption with $k = 1$ and any given $\rho > 0$. The maximum rejection probability is achieved when $\sigma_1 = \sigma_2 = \dots = \sigma_m = \rho^{-1}$ and $\sigma_{m+1} = 1$. Importantly, for any given m and $\rho > 0$, the cutoff $\underline{c}_{m,\rho}$ can be easily computed numerically, due to the monotonicity of the function $\bar{H}_m(c, \rho)$ in c . Accordingly, Theorem 4.4 gives a closed-form critical value of our t -test for any significance level less than or equal to

$$\underline{\alpha}_{m,\rho} \equiv \mathbb{P} \left[|t_{m-1}| \sqrt{\rho^2 + \frac{1}{m}} > \underline{c}_{m,\rho} \right]. \quad (17)$$

That is, for any significance level $\alpha \in (0, \underline{\alpha}_{m,\rho}]$, a valid critical value for our two-sided t -test is $\sqrt{\rho^2 + m^{-1}} t_{m-1, 1-\frac{\alpha}{2}}$, where $t_{m-1, 1-\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ quantile of the t -distribution with degree of freedom $m - 1$.

Table 2 reports the largest significance level $\underline{\alpha}_{m,\rho}$ such that our two-sided t -test has a simple closed-form expression for the critical value, under Assumption 3.2 with $k = 1$ and various values of (m, ρ) . Note that our t -test can handle all values of $m \geq 2, k \in \{1, \dots, m\}$ and $\rho \geq 0$, but it generally involves one-dimensional optimization as described in Section 4.2.3. We can conduct a similar theoretical investigation to simplify the optimization of the rejection probability under Assumption 3.2 with $k \geq 2$. However, in this case, the potential maximizers cannot be reduced to a single point, so one-dimensional numerical optimization is still required. We relegate the detailed discussion under Assumption 3.2 with $k \geq 2$ to the supplementary material.

4.4 Power of the t -test

We now investigate the power of the proposed t -test under the alternative hypothesis \bar{H}_1 in (9) with $\delta \neq 0$. Without loss of generality, we assume that $\delta > 0$. The theorem below gives a lower bound for the power of the t -test.

Theorem 4.5. Suppose that Assumption 4.1 holds for some $\delta > 0$, and define the t -statistic T_m

Table 2: Largest significance value in (17) such that our two-sided t -test has a closed-form critical value under various values of (m, ρ) .

$m \setminus \rho$	0.1	0.2	0.5	1	2	3	4	5	10
5	3.950	4.418	6.438	9.456	11.866	12.505	12.647	12.693	12.770
10	3.911	4.510	7.313	9.026	9.404	9.435	9.480	9.486	9.504
20	3.684	4.615	6.337	6.711	6.829	6.850	6.874	6.866	6.874
25	3.491	4.442	6.068	6.278	6.354	6.350	6.357	6.365	6.389
50	3.768	4.663	5.308	5.330	5.391	5.437	5.435	5.436	5.441

as in (8). Then, for any finite $c > 0$,

$$\mathbb{P}[|T_m| > c] \geq \mathbb{P}[T_m > c] \geq 1 - \frac{1}{\delta^2} \left[\sigma_{m+1}^2 + \frac{2(c^2 + m^{-1})}{m} \sum_{j=1}^m \sigma_j^2 \right].$$

The lower bound of power in Theorem 4.5 increases with δ , where δ corresponds to $\sqrt{n}(\mu_1 - \mu_0)$ in our large-sample inference for a finite number of clusters as shown in Theorem 4.1. If the gap between the means of the treated and control clusters is bounded away from zero, then, as the sample size $n \rightarrow \infty$, the power of our t -test with any finite critical value will converge to 1. In addition, Theorem 4.5 also provides a rough power estimate when we have some information about the treatment effect size and the variances for the treated and control clusters.

5 Simultaneous inference

Recall that Assumption 3.2 involves two parameters ρ and k . They represent the allowable degree of relative heterogeneity between treated and control clusters. Specifically, larger values of ρ and k correspond to a greater degree of allowable heterogeneity. In practice, specifying the values of (ρ, k) might be challenging, and it is often desirable to perform multiple analyses for a wide range of values for (ρ, k) . This raises the question of how to interpret the results of the analysis under different values of (ρ, k) . In this section, we show that the analyses over all possible values of (ρ, k) can be simultaneously valid, without the need of any adjustment due to multiple analyses. Moreover, the analysis results can be easily visualized and interpreted. Similar simultaneous inference has been used for sensitivity analysis of observational studies (Cui and Li, 2025; Wu and Li, 2025).

We first introduce several notations to denote the true relative heterogeneity of the standard deviations between treated and control clusters. For any $k \in \{1, \dots, m\}$, define

$$\rho_k^* \equiv \inf \{ \rho \geq 0 : \sigma_{m+1} \leq \rho \sigma_{(k)} \} = \begin{cases} \frac{\sigma_{m+1}}{\sigma_{(k)}}, & \text{if } \sigma_{(k)} > 0, \\ 0, & \text{if } \sigma_{(k)} = 0 \text{ and } \sigma_{m+1} = 0, \\ \infty & \text{if } \sigma_{(k)} = 0 \text{ and } \sigma_{m+1} > 0, \end{cases} \quad (18)$$

where σ_{m+1} and $\sigma_{(k)}$ in (18) denote the true standard deviation of the treated cluster and that of the control cluster at rank k . Consequently, the values of ρ_k^* reflect the true relative heterogeneity between the treated and control clusters. In particular, larger values of ρ_k^* indicate greater heterogeneity between the treated and control clusters.

We now explain the key idea for our simultaneous inference procedure. In Section 4, we test the null hypothesis in (9) of no mean difference between the treated and control clusters under Assumption 3.2 for some $k \in \{1, \dots, m\}$ and $\rho \geq 0$, which is equivalent to that $\rho_k^* \leq \rho$ with ρ_k^* defined as in (18). Importantly, we can reinterpret the t -test in Algorithm 3.1 as a valid test for the null hypothesis of $\rho_k^* \leq \rho$ about the true relative heterogeneity in (18), under the assumption of no mean difference between treated and control clusters (i.e., $\delta = 0$). By standard test inversion, we can construct confidence intervals for ρ_k^* , which must be one-sided confidence intervals with unbounded right endpoints and thus provide essentially lower bounds on the true relative heterogeneity, under the assumption that $\delta = 0$. Moreover, these confidence intervals will be simultaneously valid across all $k \in \{1, \dots, m\}$. We summarize the results in the following theorem, followed by a discussion of its practical implications and interpretation. For any $z \in \mathbb{R}$, we write $(z, \infty] \equiv (z, \infty) \cup \{\infty\}$ and analogously $[z, \infty] \equiv [z, \infty) \cup \{\infty\}$.

Theorem 5.1. *Let $\alpha \in (0, 1)$. Suppose Assumption 3.2 holds with $\delta = 0$. Let T_m be the test statistic as defined in (8).*

- (i) *For any $k \in \{1, \dots, m\}$, an $(1 - \alpha)$ -confidence set for ρ_k^* in (18) is*

$$\mathcal{I}_{m,\alpha,k} \equiv \{ \rho \geq 0 : p_m(|T_m|; k, \rho) > \alpha \} \cup \{ \infty \} = (\hat{\rho}_{m,\alpha,k}, \infty] \text{ or } [\hat{\rho}_{m,\alpha,k}, \infty], \quad (19)$$

which must be an one-sided confidence interval with $\hat{\rho}_{m,\alpha,k} \equiv \inf \mathcal{I}_{m,\alpha,k}$.

- (ii) *The confidence intervals in (i) are simultaneously valid across all $k \in \{1, \dots, m\}$, in the sense that*

$$\mathbb{P} [\rho_k^* \in \mathcal{I}_{m,\alpha,k} \text{ for all } k \in \{1, \dots, m\}] \geq 1 - \alpha.$$

The interpretation of Theorem 5.1 is as follows. If there is no mean difference between treated and control clusters (i.e., $\delta = 0$), then we need to believe that, with $(1 - \alpha)$ confidence level, the relative heterogeneity ρ_k^* in (18) must be greater than (or equal to) $\hat{\rho}_{m,\alpha,k}$, for all $k \in \{1, \dots, m\}$. That is, with $(1 - \alpha)$ confidence level, the treated standard deviation σ_{m+1} must be at least $\hat{\rho}_{m,\alpha,k}$ times the control standard deviation $\sigma_{(k)}$ at rank k , for all $k \in \{1, \dots, m\}$. If we question any of these statements on the true relative heterogeneity, then the assumption of $\delta = 0$ is likely to fail, or equivalently there is likely significant mean difference between the treated and control clusters. Obviously, the larger the values of $\{\hat{\rho}_{m,\alpha,k}\}_{k=1}^m$, the stronger the evidence for a nonzero treatment effect. In practice, we can easily visualize these confidence intervals by, say, plotting k against $\hat{\rho}_{m,\alpha,k}$; this is illustrated in the two empirical applications in Section 7 ahead.

We now discuss the computation for the confidence intervals in Theorem 5.1. By the definition of the p -value $p_m(c; k, \rho)$ in (11) and the fact that the set $\mathcal{S}_m(k, \rho)$ in (10) increases as k or ρ increase, $p_m(c; k, \rho)$ must be nondecreasing in both k and ρ . The monotonicity in ρ then explains the equivalent one-sided form of the set in (19). Thus, we can use the bisection method to find the thresholds $\{\hat{\rho}_{m,\alpha,k}\}_{k=1}^m$. In addition, the monotonicity in k implies the threshold $\hat{\rho}_{m,\alpha,k}$ is nonincreasing with k . Hence, we can first find $\hat{c}_{m,\alpha,1}$, and then sequentially use $\hat{c}_{m,\alpha,k-1}$ as an upper bound for $\hat{\rho}_{m,\alpha,k}$ in the bisection method, for $2 \leq k \leq m$.

Remark 5.1. Hagemann (2024) also reported thresholds like $\{\hat{\rho}_{m,\alpha,k}\}$ for his test. However, his test only works when $k = 1$ or 2 , while our test works for any $k \in \{1, \dots, m\}$. In addition, as shown in Theorem 5.1, these thresholds can be interpreted as lower confidence bounds for the true relative heterogeneity, and they are simultaneously valid, indicating that the confidence bounds for all k are indeed “free lunch” added to those for $k = 1$ or 2 . ■

6 Simulations

In this section, we consider two sets of numerical exercises to compare the performance of the t -test against other methods for conducting inference with a single treated cluster.

6.1 Simulation design 1: normal means

The first simulation design generates data using normal distributions. The data generating process (DGP) is as follows. We generate normal random variables as in Assump-

tion 4.1. In particular, $\{\psi_j\}_{j=1}^m$ are normally distributed with mean 0, ψ_{m+1} has mean δ , $\sigma_{m+1}^2 = \rho^2$, and $\{\sigma_j^2\}_{j=1}^m$ are specified below.

We consider two different DGPs to generate the random variables for the control clusters.

DGP 1. $\sigma_j^2 = 1$ for $j = 1, \dots, m$.

DGP 2. $\sigma_j^2 = 1 + \frac{j-1}{m-1}$ for $j = 1, \dots, m$.

DGP 1 is the homogeneous design in which all control random variables have the same variance. We introduce heterogeneity in **DGP 2** and allow the variance to vary between 1 and 2. The goal is to test the two-sided hypothesis (9), under various heterogeneity parameter ρ , cluster size m , and significance level α . We compare the performance of our t -test with Hagemann (2024) in terms of size and power. This is because both of our tests work with a single treated cluster and a finite number of control clusters, and are valid under a certain relative heterogeneity assumption.

In the simulations, we consider ρ from 0.1 to 2 with step size 0.1, $m \in \{5, 10, 25, 50\}$, and $\alpha \in \{0.01, 0.05, 0.1\}$. We focus on two-sided tests and choose ρ in our t -test and Hagemann (2024) so that it matches the one used in the DGP, i.e., the relative heterogeneity assumption is always correctly specified. The results are based on 500,000 Monte Carlo replications. Figure 3 shows the result for $\alpha = 0.05$ under **DGP 1** when $k = 1$. The figure plots the rejection rate against various values of ρ . Each facet represents a specific combination of m and δ . $\delta = 0$ corresponds to the results under the null. $\delta > 0$ corresponds to the results under the alternative. Note that Hagemann (2024)'s result does not appear or only partially appear in some of the facets. This is because Hagemann (2024) may not necessarily be able to find a weight for his rearrangement test such that his test can be shown to be valid for some combinations of m , ρ and k .

Figure 3 shows that our test performs favorably when compared to Hagemann (2024). Both of our tests control size. The t -test is more powerful, especially when ρ is small. As ρ increases above 1, the power difference decreases, but we are still more powerful.

Next, Figure 4 reports the results for **DGP 2** that includes more heterogeneity among the control random variables for $\alpha = 0.05$ and $k = 1$. As predicted by the theory, our t -test becomes more conservative when more heterogeneity are included. Our t -test continues to control size and has power against the alternative. Moreover, our test outperforms Hagemann (2024)'s test in most cases. In the supplementary material, we report the results for $k = 1$ and $k = 2$ for both DGP at various significance levels.

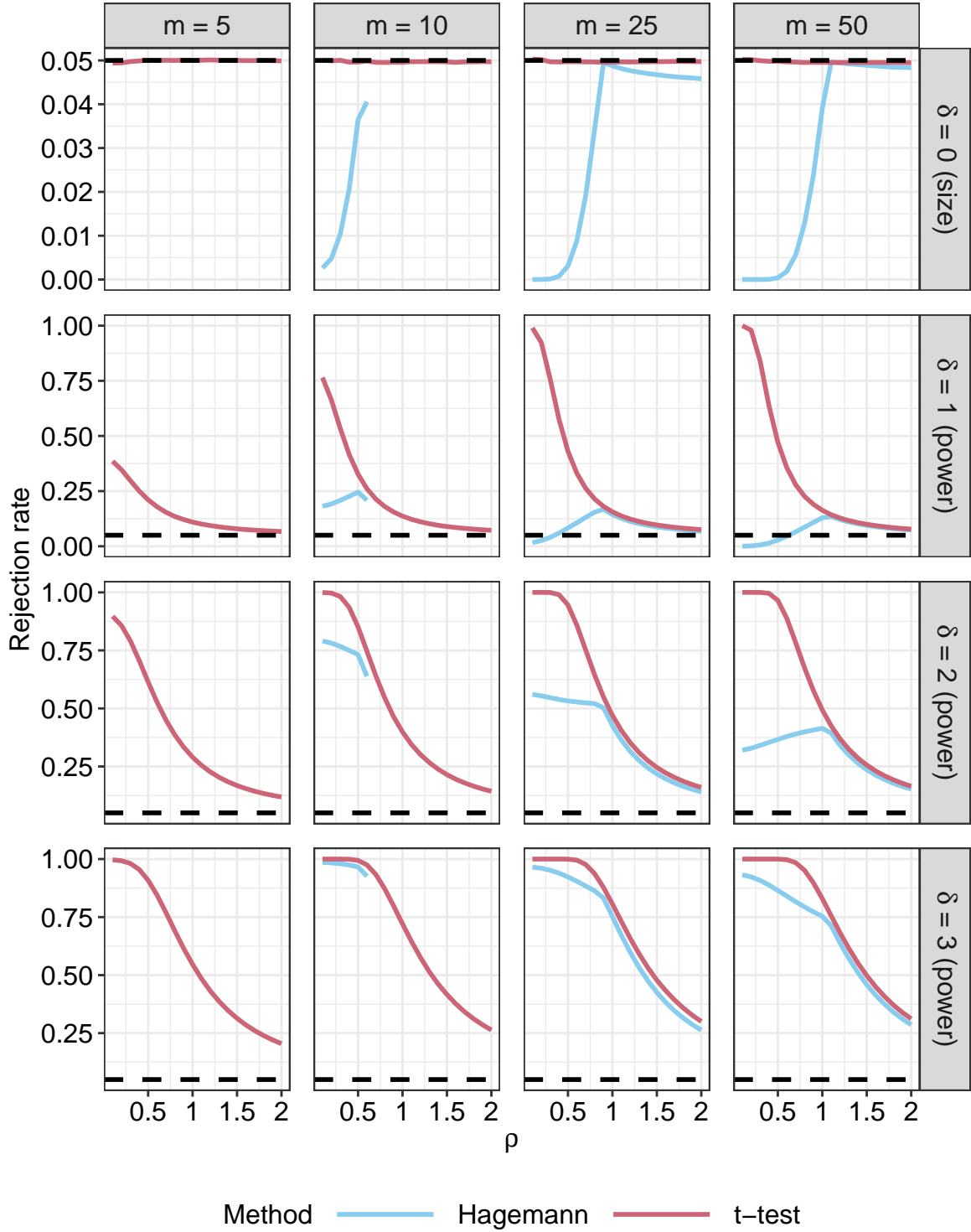


Figure 3: Probability of rejection against heterogeneity parameter ρ for various cluster size m and alternatives δ at $\alpha = 0.05$ for **DGP 1** of simulation design 1.

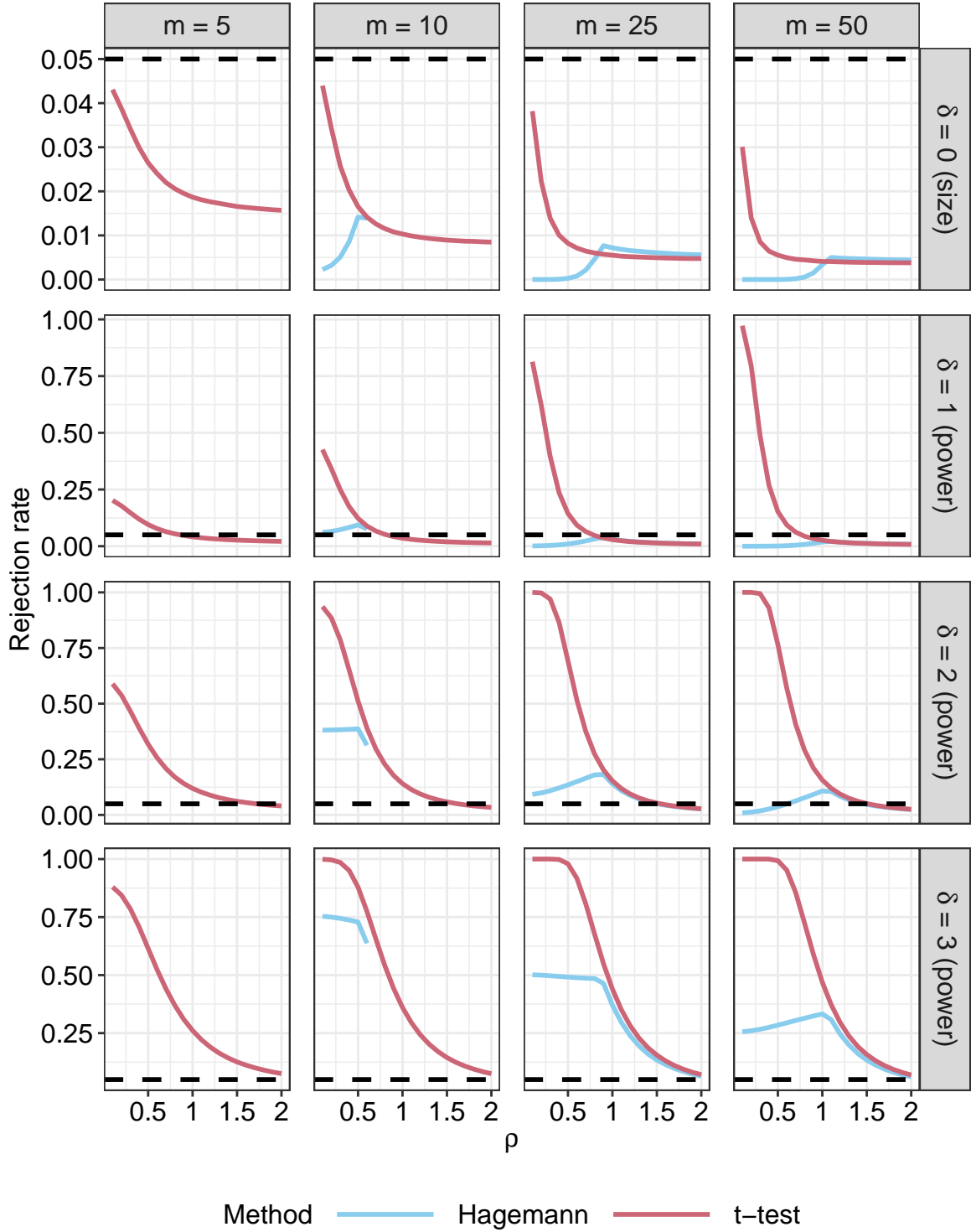


Figure 4: Probability of rejection against heterogeneity parameter ρ for various cluster size m and alternatives δ at $\alpha = 0.05$ for **DGP 2** of simulation design 1.

6.2 Simulation design 2: two-way fixed effects

Next, we conduct a simulation exercise with two-way fixed effects as in Example 2.3. We consider a design that is based on the ones used in Conley and Taber (2011) and Hagemann (2024). As before, let $\mathcal{J}_0 \equiv \{1, \dots, m\}$ be the control clusters and $\mathcal{J}_1 \equiv \{m+1\}$ be the treated cluster. Let $\mathcal{T} \equiv \{1, \dots, 10\}$ be the total number of time periods. Let $t_0 = 6$ be the intervention period and $D_{jt} \equiv \mathbb{1}[j \in \mathcal{J}_1 \text{ and } t > t_0]$ for $j \in \mathcal{J}$ and $t \in \mathcal{T}$. Each simulated data is generated from the following two-way fixed effects model

$$Y_{jt} = \alpha_t + \gamma_j + \theta D_{jt} + U_{jt}, \quad (20)$$

where $\alpha_t = 1$ for all $t \in \mathcal{T}$ and $\gamma_j = 2\mathbb{1}[j \leq \frac{m}{2}] - 1$ for all $j \in \mathcal{J}$. We test the null of $\theta = 0$ and consider $\theta \in \{0, 1, 2, 3\}$ when generating the data. We consider the following DGPs based on model (20). The first four DGPs are the same as the ones considered in Hagemann (2024). The remaining two DGPs are based on the error distributions considered in Conley and Taber (2011).

DGP 1. Generate $U_{jt} = \eta U_{j,t-1} + \sigma^{\mathbb{1}[j=m+1]} V_{jt}$ where $\eta = 0.5$ and V_{jt} are independently distributed standard normal random variables.

DGP 2. Same as DGP 1, but use $\eta = 0.1$.

DGP 3. Same as DGP 1, but use $\eta = 0.9$.

DGP 4. Same as DGP 1, but V_{jt} follows a normalized χ_2^2 distribution with mean 0 and variance 1.

DGP 5. Same as DGP 1, but use $V_{jt} \sim \text{Uniform}[-\sqrt{3}, \sqrt{3}]$.

For each DGP, we are interested in testing the two-sided hypothesis of

$$H_0 : \theta = 0 \quad \text{vs} \quad H_1 : \theta \neq 0.$$

In addition, we consider simulations with $m \in \{5, 10, 25, 50\}$, $\sigma \in \{0.1, 0.5, 1, 2\}$, and $\alpha = 0.05$. We report the rejection rate curves based on 5,000 Monte Carlo simulations. For each DGP, we consider conducting inference using the following four methods:

t The t -test proposed in our paper.

H The rearrangement test by Hagemann (2024).

CT The procedure of Conley and Taber (2011) that assumes homogeneity.

FP The bootstrap procedure of Ferman and Pinto (2019).

Figures 5 and 6 show the results for DGPs 1 and 2 under $\alpha = 0.05$ across different numbers of clusters m and true relative heterogeneity σ . We set the relative heterogeneity parameter ρ for both our and Hagemann (2024)'s method in all the DGPs to be the corresponding σ , so that the relative heterogeneity assumption is correctly specified. We also incorporate the correct value of σ in applying FP to specify the heterogeneity.

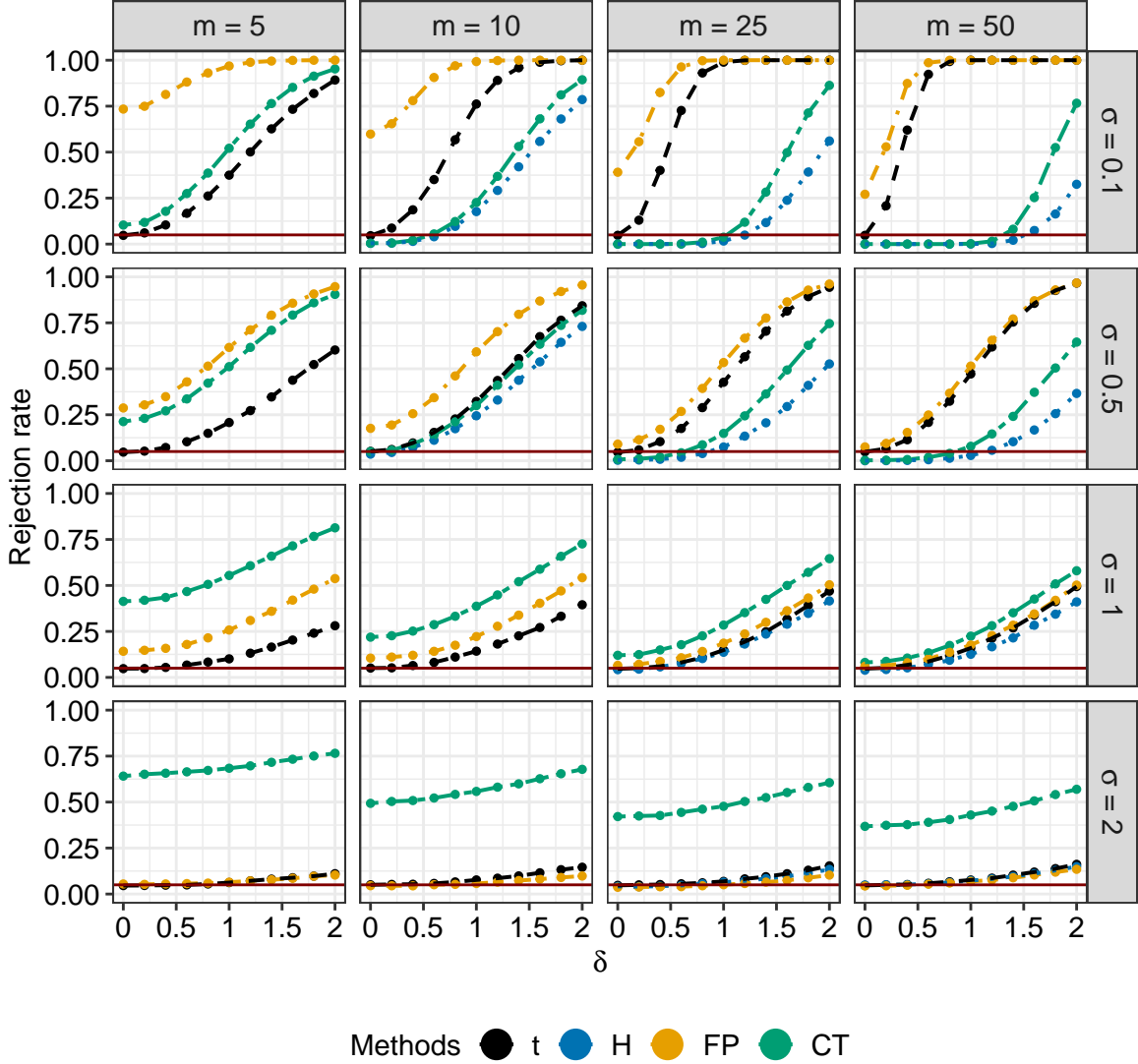


Figure 5: Simulation results for **DGP 1** of simulation design 2 at $\alpha = 0.05$.

It can be seen that our t -test performs well and compares favorably to other methods. In particular, the CT test exhibits inflated Type I error rates when the number of control clusters m is small and σ is large. This inflation arises because the validity of CT relies on the assumption of an infinite number of clusters and homogeneous variances between

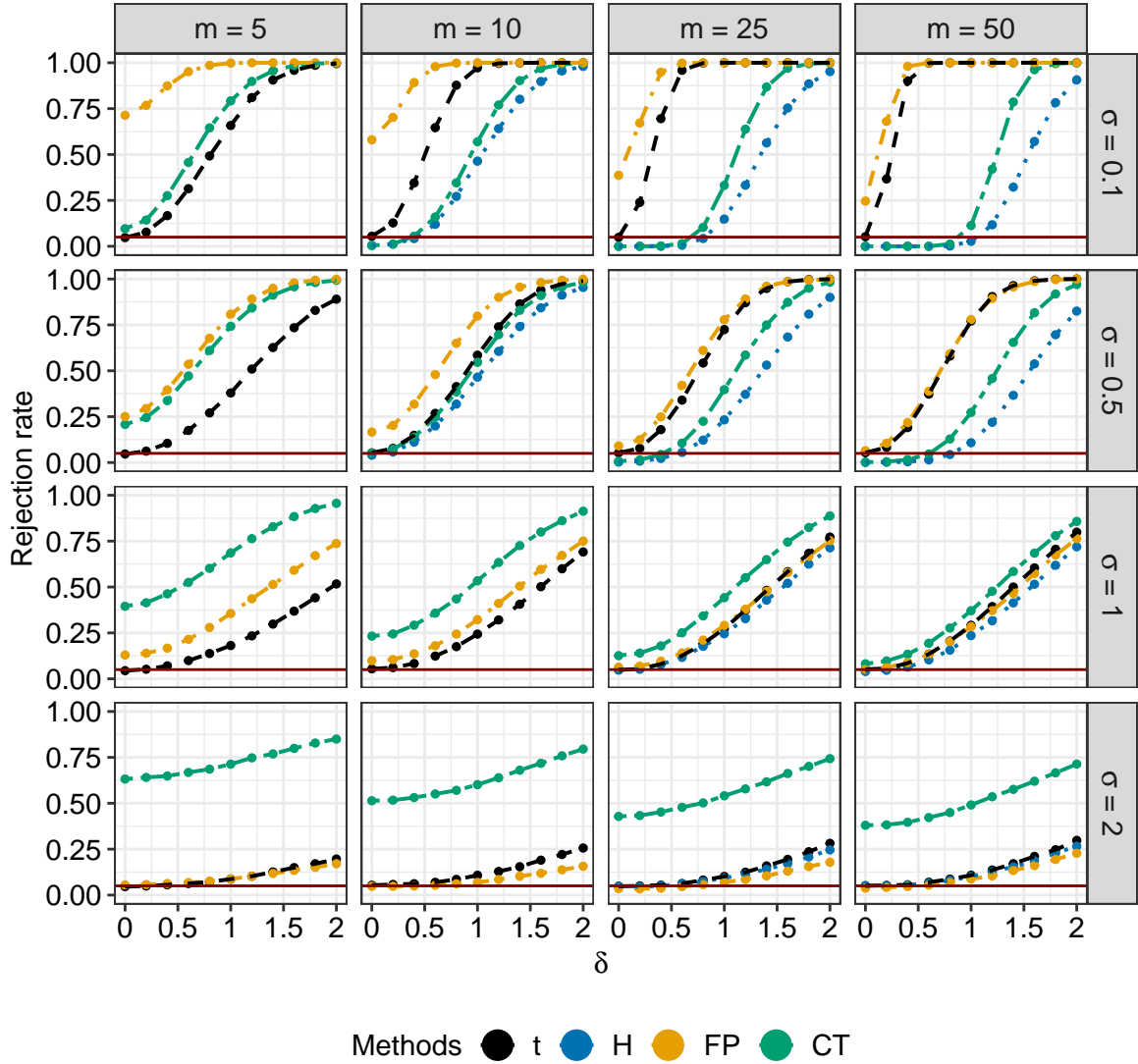


Figure 6: Simulation results for **DGP 2** of simulation design 2 at $\alpha = 0.05$.

the treated and control clusters. The FP test also shows an inflated Type I error when m is small, with the inflation becoming more pronounced as σ decreases. This again reflects its reliance on the asymptotic validity under an infinite number of clusters. Notably, we assume that FP has access to the true heterogeneity between the treated cluster and each control cluster, which is a stronger assumption than those required by our t -test or [Hagemann \(2024\)](#)'s rearrangement test. Both our test and that of [Hagemann \(2024\)](#) successfully control Type I error. In contrast, our test is applicable regardless of the number of control clusters and demonstrates higher power across different levels of heterogeneity. The results for DGPs 3 to 5 are contained in the supplementary material.

7 Empirical applications

In this section, we illustrate the t -test proposed in this paper with two recent empirical applications. For each of the empirical applications, we report two sets of results. First, we try to find the largest ρ such that the null hypothesis is rejected when $k = 1$ among various regression specifications of the empirical examples. Among those specifications where the null can be rejected at some ρ , we conduct simultaneous inference as discussed in Section 5.

7.1 Empirical application 1: [Depew and Swensen \(2022\)](#)

[Depew and Swensen \(2022\)](#) examines the impact of the 1911 New York State Sullivan Act on mortality rate. The Sullivan Act required New York citizens to obtain a permit and license to carry concealable weapon. We consider the following two-way fixed effects model:

$$\text{Outcome}_{jt} = \beta \text{Treated}_{jt} + \alpha_t + \gamma_j + \epsilon_{jt}, \quad (21)$$

where Treated_{jt} equals 1 if state j is New York and t is the post-treatment period, α_t is the year fixed effects, γ_j is the state fixed effects, and ϵ_{jt} is the residual term. They consider the following four outcome variables: homicide rate, suicide rate, gun suicide rate, and non-gun suicide rate.

They report cluster-robust standard errors and p -value from the Wild cluster bootstrap. There is only nine control clusters in this empirical application. Hence, [Hagemann \(2024\)](#)'s test may not be applicable because there can be no weights such that his rearrangement test is valid for some relative heterogeneity parameters ρ .

We are interested in testing

$$H_0 : \beta = 0 \quad \text{vs} \quad H_1 : \beta \neq 0. \quad (22)$$

Table 3 summarizes the estimation and inference results of the baseline model described in (21) for the four outcomes described above. The point estimates and the Wild cluster bootstrap p -values are taken from Table 2 of [Depew and Swensen \(2022\)](#).⁸

The remaining rows of Table 3 report the largest ρ such that the null can be rejected for significance levels $\alpha = 0.01, 0.05$, and 0.1 . For entries with "NA" in the table, it refers to the situation where there does not exist a $\rho \geq 0$ such that the null is rejected. If we

⁸The point estimates and p -values that [Depew and Swensen \(2022\)](#) report are based on weighted least squares.

Table 3: Regression and inference results for empirical application 1 with $k = 1$.

	Homicide rate	Suicide rate	Gun suicide rate	Non-gun suicide rate
Point estimate	-0.04	-0.85	-1.42	0.54
Wild p -value	0.87	0.29	0	0.28
Rearrangement test (Hagemann, 2024)				
$\alpha = 0.01$	NA	NA	NA	NA
$\alpha = 0.05$	NA	NA	NA	NA
$\alpha = 0.1$	NA	NA	NA	NA
t-test				
$\alpha = 0.01$	NA	NA	0.65	NA
$\alpha = 0.05$	NA	NA	1.03	NA
$\alpha = 0.1$	NA	NA	1.3	NA

focus on the gun suicide rate under $\alpha = 0.1$, it means that the null can be rejected when the variance in New York is at most $1.3^2 \approx 1.69$ larger than the smallest variance in the control clusters. The true relative level of heterogeneity such that the null can be rejected may be related to the population size of the various states. In particular, New York state has a larger population than the other control states in 1911 ([Federal Reserve Economic Data, 2025](#)): the population of New York state was 9.249 million, and the largest control state was Massachusetts with 3.383 million and the smallest control state was Vermont with 0.358 million.

Next, we conduct simultaneous inference for our t -test as in Section 5. We focus on the outcome “gun suicide rate” as we can find ρ such that the null is rejected when $k = 1$. Figure 7 reports the results for $\alpha \in \{0.01, 0.05, 0.1\}$ on the range of ρ for various k such that the null can be rejected. The shaded area represents the simultaneous confidence region for true relative heterogeneity when there is no treatment effect.

7.2 Empirical application 2: [Hiraiwa et al. \(2024\)](#)

[Hiraiwa et al. \(2024\)](#) examine whether firms enforce noncompete agreements (NCAs). NCAs are restrictions that prevent workers from joining or starting competing firms. In 2020, Washington state passed a law that prohibits NCAs for workers earning below a certain threshold. The threshold was \$100,000 per year in 2020 and \$101,390 per year in

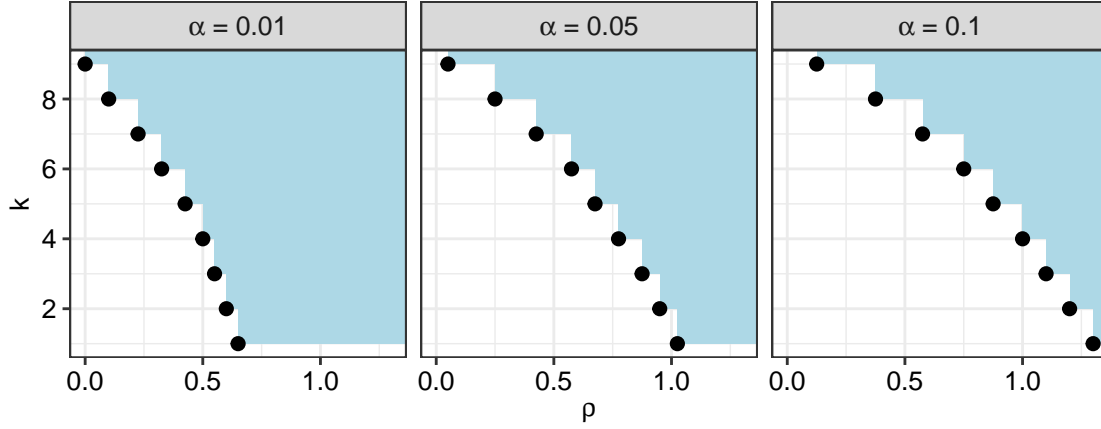


Figure 7: Simultaneous inference for empirical application 1 at different significance levels α .

2021. They study the impact of being above or below the threshold on employment for different earnings bins. They consider the following two-way fixed effects model in their panel B of Table 1:

$$\log \text{Emp}_{b,t} = \beta \text{Treated}_{b,t} + \alpha_t + \gamma_b + \epsilon_{b,t},$$

where $\text{Emp}_{b,t}$ is the employment count of bin b at year t , $\text{Treated}_{b,t}$ equals 1 for the focal bin in the focal year, α_t and γ_b are the year and bin fixed effects, and $\epsilon_{b,t}$ is the error term. They have 30 clusters (income bins). They use one-sided randomization inference and found that there is no significant effect.

We apply our method and [Hagemann \(2024\)](#). We are interested in testing the two-sided hypothesis as in (22). Table 4 summarizes our results. Each column contains the result for a specific focal year and definition of the treatment variable. For instance, column 2 refers to focal year 2020, and the treatment variable is defined to be equal to 1 when the income bin is just above the threshold, i.e., \$100-101.389k. The remaining columns are defined similarly. For each regression specification, we find the largest ρ such that the null is rejected. For entries with “NA” in the table, it refers to the situation where there does not exist a $\rho \geq 0$ such that the null is rejected.

Next, we conduct simultaneous inference for our t -test as in the first empirical application. We focus on the two variables for focal year 2021 because we can find ρ such that the null is rejected when $k = 1$. In particular, we report the range of ρ for various k as long as there exists ρ such that the null can be rejected. Figure 8 reports the results for $\alpha \in \{0.01, 0.05, 0.1\}$.

Table 4: Regression and inference results for empirical application 2 with $k = 1$.

	Focal year 2020		Focal year 2021	
	Above threshold \$100–101.389k	Below threshold \$98.61–100k	Above threshold \$101.39–102.78k	Below threshold \$100–101.39k
Point estimate	-0.001	-0.002	-0.038	-0.024
Rearrangement test (Hagemann, 2024)				
$\alpha = 0.01$	NA	NA	NA	NA
$\alpha = 0.05$	NA	NA	NA	NA
$\alpha = 0.1$	NA	NA	NA	NA
<i>t</i>-test				
$\alpha = 0.01$	NA	NA	0.31	0.14
$\alpha = 0.05$	NA	NA	0.45	0.25
$\alpha = 0.1$	NA	NA	0.56	0.33

8 Conclusion

In this paper, we propose a t -test to conduct inference with a single treated cluster under a certain relative heterogeneity assumption. The t -statistic and the associated critical values are easy to calculate in many empirically relevant applications. We show that our test performs favorably when compared to other methods in the literature. We also show that our test is valid with weaker assumptions.

A Appendix for the main text

A.1 Lemmas for Section 4.2.3

We first state the lemma that the maximum rejection probability in (11) must be obtained at some finite values of $\{\sigma_j\}_{j=1}^{m+1}$. This implies that we do not need to worry about the boundary case where some of $\{\sigma_j\}_{j=1}^{m+1}$ approach infinity.

Lemma A.1. *For any $k \in \{1, \dots, m\}$ and $c \neq m^{-1/2}$,⁹ the maximum rejection probability*

⁹Similar to the footnote for Theorem 4.1 and as discussed in Remark 4.4, we will consider values of c greater than $m^{-1/2}$ for most conventional significance levels.

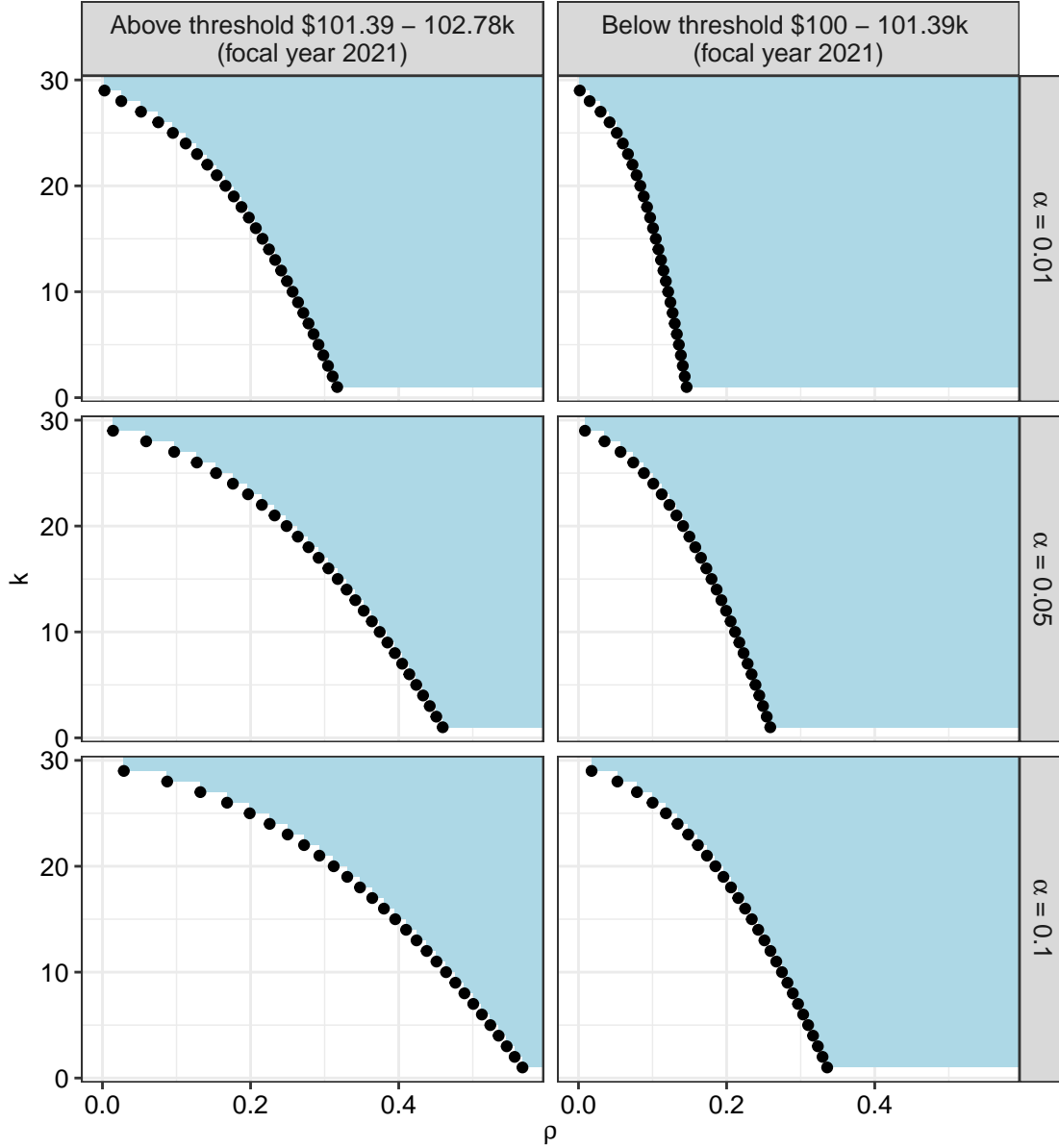


Figure 8: Simultaneous inference for empirical application 2 for different outcomes and significance levels α .

$p_m(c; k, \rho)$ in (11) must be obtained at some $(\sigma_1, \dots, \sigma_m, \sigma_{m+1}) \in \mathcal{S}_m(k, \rho)$.

Next, we investigate possible maximizers for the rejection probability. From Section 4.2.1, when $\sigma_{m+1} = 0$, the rejection probability $\mathbb{P}_0[|T_m| > c]$ of our t -test is at most $p_{m,0}(c)$ defined as in (12), which can be achieved at some values of $\{\sigma_j\}_{j=1}^m$ that satisfy our relative heterogeneity assumption, regardless of the values of ρ and k . Thus, in the following, we consider only the case where $\sigma_{m+1} > 0$.

Recall the integral representation of the rejection probability in Lemma 4.1. Our goal is to optimize (13) over $\gamma_j \equiv \frac{\sigma_j}{\sigma_{m+1}}$, $j = 1, \dots, m$, under the constraint imposed by the relative heterogeneity assumption. Such an optimization can be numerically challenging even for moderate m . The following is a key lemma that can greatly reduce the set of possible maximizers for the rejection probability.

Lemma A.2. *Consider the integral representation in (13) for the rejection probability. Fix the values of $c > 0$, $\gamma_3, \dots, \gamma_m \geq 0$ and $\theta_{m+1} < 0$, and view the rejection probability $\mathbb{P}_0[|T_m| > c] \equiv \bar{p}_m(c; \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_m)$ as a function of only γ_1 , where γ_2 is uniquely determined by $c, \gamma_1, \gamma_3, \dots, \gamma_m$ and θ_{m+1} . If $c \neq m^{-1/2}$, γ_1 and γ_2 are both positive and they are not equal, i.e., $\gamma_1 \neq \gamma_2$, and the first derivative of $\bar{p}_m(c; \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_m)$ over γ_1^2 is zero, then the second derivative of $\bar{p}_m(c; \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_m)$ over γ_1^2 must be positive.*

Lemma A.2 has an important implication. Specifically, given values of $c, \gamma_3, \dots, \gamma_m$ and θ_{m+1} , if γ_1 is in the interior of the feasible region for optimization, and $\gamma_1 \neq \gamma_2$, then $(\gamma_1, \gamma_2, \dots, \gamma_m)$ cannot be a maximizer, or even a local maximizer, for the rejection probability $\bar{p}_m(c; \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_m)$. Moreover, because T_m is invariant to permutations of control clusters, the value of $\bar{p}_m(c; \gamma_1, \gamma_2, \dots, \gamma_m)$ is invariant to permutations of $(\gamma_1, \dots, \gamma_m)$. Therefore, Lemma A.2 essentially applies to any pair (γ_i, γ_j) for $i \neq j$.

For any $k \in \{1, \dots, m\}$ and $\rho > 0$, suppose now that $(\sigma_1, \dots, \sigma_m, \sigma_{m+1})$ is a maximizer for the maximum rejection probability in (11), and $\sigma_{m+1} > 0$. Then $(\gamma_1, \dots, \gamma_m) = (\frac{\sigma_1}{\sigma_{m+1}}, \dots, \frac{\sigma_m}{\sigma_{m+1}})$ must be a maximizer of $\bar{p}_m(c; \gamma_1, \dots, \gamma_m)$ over all $(\gamma_1, \dots, \gamma_m) \in \mathbb{R}_{\geq 0}^m$ satisfying that $\gamma_{(k)} \geq \rho^{-1}$, where $\gamma_{(k)}$ is the k th smallest value of $\{\gamma_j\}_{j=1}^m$. From Lemma A.2, this cannot be true if there exists $i \neq j \in \{1, 2, \dots, m\}$ such that (i) $\gamma_i \neq \gamma_j$, and (ii) both γ_i and γ_j are not in the set $\{0, \rho^{-1}\}$, where the latter ensures that γ_i and γ_j are not in the boundary of the feasible region so that we can apply Lemma A.2. Therefore, to maximize the rejection probability, it suffices to consider only the cases where γ_j , for $1 \leq j \leq m$, is either 0, or ρ^{-1} , or some common value in $\mathbb{R}_{\geq 0}$. More specifically, under Assumption 3.2 for any given k and $\rho > 0$, it suffices to consider the following possible values for $(\gamma_1, \dots, \gamma_m)$ for $0 \leq m_0 \leq k-1$ and $0 \leq m_1 \leq m - m_0$:

$$\begin{pmatrix} \rho^{-1} \mathbf{1}_{m_1} \\ \mathbf{0}_{m_0} \\ \gamma \mathbf{1}_{m-m_1-m_0} \end{pmatrix}^\top, \quad \text{where } \gamma \in \begin{cases} \mathbb{R}_{\geq 0}, & \text{if } m_1 \geq m - k + 1, \\ [\rho^{-1}, \infty), & \text{if } m_1 < m - k + 1. \end{cases} \quad (23)$$

In (23), we essentially enumerate the possible numbers of $\{\gamma_j\}_{j=1}^m$ that are zero, ρ^{-1} , and some common value γ , respectively. Note that under Assumption 3.2 with given $\rho > 0$

and k , the number of terms in $\{\gamma_j\}_{j=1}^m$ that are zero is at most $(k - 1)$, and the possible range of the common value γ depends on the number m_1 of the terms in $\{\gamma_j\}_{j=1}^m$ that are already ρ^{-1} . In particular, if $m_1 < m - k + 1$, then the common value γ is at least ρ^{-1} .

A.2 Lemmas for the $k = 1$ case in Section 4.3

In this subsection, we show that the optimization in (16) can be further simplified, by excluding values of γ that cannot maximize $\bar{p}_m(c; \rho, \gamma; m_1, m_0)$ in (15). This relies crucially on the following lemma; we present a concise version below, and relegate the details to the supplementary material.

Lemma A.3. *Consider the integral representation in (13). Fix the values of $c > 0$, $m \geq 3$, $\gamma_3, \dots, \gamma_m \geq 0$ and $\theta_{m+1} < 0$, and view the rejection probability $\mathbb{P}_0[|T_m| > c] = \bar{p}_m(c; \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_m)$ as a function of only γ_1 , where γ_2 is uniquely determined by $c, \gamma_1, \gamma_3, \dots, \gamma_m$ and θ_{m+1} . If $\gamma_1 = \max_{1 \leq i \leq m} \gamma_i > \gamma_2 > 0$, $c \geq \sqrt{\frac{2(m-1)}{m(m-2)}}$, and $H_m(c; \gamma_1, \dots, \gamma_m) < 0$ ¹⁰, then the first order derivative of $\bar{p}_m(c; \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_m)$ over γ_1^2 must be negative.*

Lemma A.3 has an important implication. Specifically, given values of $\gamma_3, \dots, \gamma_m$ and θ_{m+1} , if, for some small $\epsilon > 0$, $[\gamma_1 - \epsilon, \gamma_1]$ is in the feasible region for optimization, and the conditions in Lemma A.3 holds, then $(\gamma_1, \gamma_2, \dots, \gamma_m)$ cannot be a maximizer, or even a local maximizer, for the rejection probability $\bar{p}_m(c; \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_m)$. This is because we can strictly increase this rejection probability by slightly decreasing γ_1 . Analogous to the discussion after Lemma A.2, because the rejection probability is invariant to permutations of $(\gamma_1, \dots, \gamma_m)$, we can always set $\gamma_1 = \max_{1 \leq i \leq m} \gamma_i$ without loss of generality, and apply Lemma A.3 to any γ_j with $2 \leq j \leq m$. As long as the conditions in Lemma A.3 hold for some $2 \leq j \leq m$, we can exclude the corresponding $(\gamma_1, \dots, \gamma_m)$ from the set of possible maximizers for the rejection probability.

We apply Lemma A.3 to the optimization in (16) to exclude some values of $\gamma > \rho^{-1}$ that cannot maximize $\bar{p}_m(c; \rho, \gamma; m_1, m_0)$ ¹¹. In the ideal case, we can exclude all values of $\gamma > \rho^{-1}$, and the optimization in (16) is either solved or needs only to be maximized over $\gamma \in [0, \rho^{-1}]$. In Theorem 4.4, we focus on the relative heterogeneity assumption with $k = 1$, i.e., the treated standard deviation is smaller than or equal to ρ times each of the control standard deviations. We relegate the discussion under Assumption 3.2 with $k \geq 2$ to the supplementary material.

¹⁰We give the detailed expression of $H_m(\cdot)$ in the supplementary material

¹¹Specifically, when using Lemma A.3, we set γ_1 to be γ and γ_2 to be ρ^{-1} .

A.3 t -test with a large number of clusters

Despite that our inference procedure focuses on the case with a finite number of control clusters, we briefly explore the asymptotic regime where the number of control clusters tends to infinity. This offers useful insights even when the number of control clusters is finite. Below we first state a rather weak regularity condition. We then present the main theorem and discuss its implications.

Condition A.1. As $m \rightarrow \infty$, both σ_{m+1}^2 and $m^{-1} \sum_{j=1}^m \sigma_j^2$ are bounded away from zero and infinity, and $m^{-2} \sum_{j=1}^m \sigma_j^4 \rightarrow 0$.

Theorem A.1. Under Assumption 4.1 and Condition A.1, as $m \rightarrow \infty$, we have

$$\sup_{c \in \mathbb{R}} \left| \mathbb{P}[T_m \leq c] - \mathbb{P}\left[\frac{\sigma_{m+1}\varepsilon + \delta}{\sqrt{m^{-1} \sum_{j=1}^m \sigma_j^2}} \leq c\right] \right| \rightarrow 0.$$

where $\varepsilon \sim \mathcal{N}(0, 1)$.

From Theorem A.1, when m is large, the distribution of the test statistic T_m under both the null and the alternative hypotheses can be approximated by an Gaussian distribution. This has important implications, as detailed below.

First, suppose the null hypothesis \bar{H}_0 in (9) holds (i.e., $\delta = 0$). Then, when the number of clusters is large, the rejection probability $\mathbb{P}_0[|T_m| > c]$ depends only the ratio between the treated-cluster variance σ_{m+1}^2 and the average of the control-cluster variances $m^{-1} \sum_{j=1}^m \sigma_j^2$. It is increasing in this ratio. In particular, under Assumption 3.2 for some $\rho > 0$, we have the following for any $c > 0$:

$$\mathbb{P}_0[|T_m| > c] \approx \mathbb{P}\left[\frac{\sigma_{m+1}|\varepsilon|}{\sqrt{m^{-1} \sum_{j=1}^m \sigma_j^2}} > c\right] \leq \mathbb{P}\left[\sqrt{\frac{m\rho^2}{m-k+1}} \cdot |\varepsilon| > c\right],$$

where the last inequality becomes equality when $\sigma_{m+1} = 1$, $(k-1)$ of $\{\sigma_j\}_{j=1}^m$ equal 0, and the remaining $(m-k+1)$ equal ρ^{-1} . It also implies that the critical value $\text{cv}_{m,\alpha,k,\rho}$ of a level- α test is approximately $\sqrt{\frac{m}{m-k+1}} \cdot \rho z_{\frac{\alpha}{2}}$, where $z_{\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$ upper quantile of the standard normal distribution.

Second, under the alternative hypothesis with, say, $\delta > 0$, the above discussion also

implies that the power of the level- α t -test is approximately

$$\begin{aligned}
\mathbb{P}[|T_m| > \text{cv}_{m,\alpha,k,\rho}] &\approx \mathbb{P}\left[\left|\frac{\sigma_{m+1}\varepsilon + \delta}{\sqrt{m^{-1}\sum_{j=1}^m\sigma_j^2}}\right| > \sqrt{\frac{m\rho^2}{m-k+1}}z_{\frac{\alpha}{2}}\right] \\
&= \mathbb{P}\left[\left|\varepsilon + \frac{\delta}{\sigma_{m+1}}\right| > R_{m,k,\rho}z_{\frac{\alpha}{2}}\right] \\
&\geq \mathbb{P}\left[\varepsilon + \frac{\delta}{\sigma_{m+1}} > R_{m,k,\rho}z_{\frac{\alpha}{2}}\right], \tag{24}
\end{aligned}$$

where $R_{m,k,\rho}^2$ is the ratio between the maximum possible value of $\frac{\sigma_{m+1}^2}{m^{-1}\sum_{j=1}^m\sigma_j^2}$ and the corresponding true value, and $R_{m,k,\rho}$ is the square root of $R_{m,k,\rho}^2$. From (24), the lower bound of the approximate power in (24), which is also approximately the power of an one-sided level- $\frac{\alpha}{2}$ test, is increasing in $\frac{\delta}{\sigma_{m+1}}$ and decreasing in $R_{m,k,\rho}$. This is intuitive, since $\frac{\delta}{\sigma_{m+1}}$ represents the signal-to-noise ratio, whereas $R_{m,k,\rho}$ represents how conservative we are when making the relative heterogeneity assumption.

We now give a remark that is helpful for our numerical search of the critical value when m is finite.

Remark A.1. Recall that $\gamma_j \equiv \frac{\sigma_j}{\sigma_{m+1}}$ for $1 \leq j \leq m$. From the discussion before, the maximum rejection probability is achieved when $k-1$ of $\{\gamma_j\}_{j=1}^m$ equal 0 and the remaining $(m-k+1)$ equal 1. Thus, given any significance level, we suggest using this configuration of $\{\gamma_j\}_{j=1}^m$ as an initial guess for the desired critical value for any finite m , and then use the optimization in Theorem 4.3 to verify whether such an initial guess of critical value leads to the t -test with the desired significance level; see Remark 4.5. Moreover, as demonstrated in Section 4.3, under the relative heterogeneity assumption with $k=1$, the maximizer for the rejection probability from the asymptotic analysis with large m is also valid when the number of control clusters is finite and the significance level is sufficiently small. Our numerical experiment shows that this holds more generally; at usual significance levels, the critical value from the asymptotic analysis with large m is often valid even when m is finite, for general relative heterogeneity assumption with $k \geq 1$. ■

References

ALPERT, A., S. DYKSTRA, AND M. JACOBSON (2024): “Hassle Costs versus Information: How Do Prescription Drug Monitoring Programs Reduce Opioid Prescribing?” American Economic

- Journal: Economic Policy, 16, 87–123.
- ALVAREZ, L. AND B. FERNAN (2023): “Inference in Difference-in-Differences with Few Treated Units and Spatial Correlation,” .
- ALVAREZ, L., B. FERNAN, AND K. WÜTHRICH (2025): “Inference with few treated units,” arXiv preprint arXiv:2504.19841.
- BAKER, A., B. CALLAWAY, S. CUNNINGHAM, A. GOODMAN-BACON, AND P. H. C. SANT’ANNA (2025): “Difference-in-Differences Designs: A Practitioner’s Guide,” .
- BAKIROV, N. (1989): “An extremal property of the student distribution,” Journal of Mathematical Sciences, 44, 433–440.
- BAKIROV, N. K. (1998): “Nonhomogeneous samples in the Behrens-Fisher problem,” Journal of Mathematical Sciences, 89, 1460–1467.
- BAKIROV, N. K. AND G. J. SZÉKELY (2006): “Student’s t-test for Gaussian scale mixtures,” Journal of Mathematical Sciences, 139, 6497–6505.
- BESTER, A., T. CONLEY, AND C. HANSEN (2011): “Inference with dependent data using cluster covariance estimators,” Journal of Econometrics, 165, 137–151.
- CAMERON, A., J. GELBACH, AND D. MILLER (2008): “Bootstrap-Based Improvements for Inference with Clustered Errors,” The Review of Economics and Statistics, 90, 414–427.
- CAMERON, A. C. AND D. L. MILLER (2015): “A practitioner’s guide to cluster-robust inference,” Journal of human resources, 50, 317–372.
- CANAY, I. A., J. P. ROMANO, AND A. M. SHAIKH (2017): “Randomization Tests under an Approximate Symmetry Assumption,” Econometrica, 85, 1013–1030.
- CANAY, I. A., A. SANTOS, AND A. M. SHAIKH (2021): “The Wild Bootstrap with a “small” number of “large” Clusters,” The Review of Economics and Statistics, 2021.
- CONLEY, T., S. GONÇALVES, AND C. HANSEN (2018): “Inference with Dependent Data in Accounting and Finance Applications,” Journal of Accounting Research, 56, 1139–1203.
- CONLEY, T. G. AND C. R. TABER (2011): “Inference with “Difference in Differences” with a Small Number of Policy Changes,” The Review of Economics and Statistics, 93, 113–125.
- CUI, H. AND X. LI (2025): “Robust Sensitivity Analysis via Augmented Percentile Bootstrap under Simultaneous Violations of Unconfoundedness and Overlap,” arXiv preprint arXiv:2509.13169.
- DE CHAISEMARTIN, C. AND X. D’HAULTFŒUILLE (2022): “Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey,” The Econometrics Journal, 26, C1–C30.
- DEPEW, B. AND I. SWENSEN (2022): “The Effect of Concealed-Carry and Handgun Restrictions on Gun-Related Deaths: Evidence from the Sullivan Act of 1911,” The Economic Journal, 132, 2118–2140.

- DILLENDER, M., L. JINKS, AND A. T. LO SASSO (2023): “When (and why) providers do not respond to changes in reimbursement rates,” Journal of Public Economics, 217, 104781.
- FEDERAL RESERVE ECONOMIC DATA (2025): “Resident Population by State, Annual,” <https://fred.stlouisfed.org/release/tables?rid=118&eid=259194&od=1911-01-01#>, accessed: 2025-05-25.
- FERMAN, B. AND C. PINTO (2019): “Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity,” The Review of Economics and Statistics, 101, 452–467.
- HAGEMANN, A. (2022): “Permutation inference with a finite number of heterogeneous clusters,” Working Paper.
- (2024): “Inference with a single treated cluster,” Working Paper.
- HARRIS, D. N. AND M. F. LARSEN (2023): “Taken by storm: The effects of Hurricane Katrina on medium-term student outcomes in New Orleans,” Journal of Human Resources, 58, 1608–1643.
- HIRAIWA, T., M. LIPSITZ, AND E. STARR (2024): “Do Firms Value Court Enforceability of Non-compete Agreements? A Revealed Preference Approach,” The Review of Economics and Statistics, 1–47.
- HOROWITZ, J. AND C. MANSKI (1995): “Identification and Robustness with Contaminated and Corrupted Data,” Econometrica, 63, 281–302.
- IBRAGIMOV, R. AND U. K. MÜLLER (2010): “t-Statistic Based Correlation and Heterogeneity Robust Inference,” Journal of Business & Economic Statistics, 28, 453–468.
- (2016): “Inference with Few Heterogeneous Clusters,” The Review of Economics and Statistics, 98, 83–96.
- KLINE, P. AND A. SANTOS (2013): “Sensitivity to missing data assumptions: Theory and an evaluation of the U.S. wage structure,” Quantitative Economics, 4, 231–267.
- KUMAR, A. AND C.-Y. LIANG (2024): “Labor Market Effects of Credit Constraints: Evidence from a Natural Experiment,” American Economic Journal: Economic Policy, 16, 1–26.
- LAU, C. P. (2025): “Combining Clusters for the Approximate Randomization Test,” Working Paper.
- MACKINNON, J. G., M. ØRREGAARD NIELSEN, AND M. D. WEBB (2023): “Cluster-robust inference: A guide to empirical practice,” Journal of Econometrics, 232, 272–299.
- MAKSHANOV, A. AND O. SHALAEVSKII (1977): “Some problems of asymptotic approximations of distributions,” Zapiski Nauchnykh Seminarov POMI, 74, 118–138.
- MASTEN, M. A. AND A. POIRIER (2020): “Inference on breakdown frontiers,” Quantitative Economics, 11, 41–111.
- OLDEN, A. AND J. MØEN (2022): “The triple difference estimator,” The Econometrics Journal, 25, 531–553.

- ROTH, J., P. H. SANT'ANNA, A. BILINSKI, AND J. POE (2023): "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," Journal of Econometrics, 235, 2218–2244.
- WANG, J. AND K. BURKE (2022): "The effects of disclosure and enforcement on payday lending in Texas," Journal of Financial Economics, 145, 489–507.
- WANG, Y. AND X. LI (2022): "Rerandomization with diminishing covariate imbalance and diverging number of covariates," The Annals of Statistics, 50, 3439 – 3465.
- WU, D. AND X. LI (2025): "Sensitivity analysis for quantiles of hidden biases in matched observational studies," Journal of the American Statistical Association, 1–12.

Supplementary material for “Cluster-robust inference with a single treated cluster using the t -test”

S1 Technical lemmas for the t -test

Throughout this section, we will study t -test based on independent normal observations under Assumption 4.1. In Section S1.1, we will show that the t -test reject the null if and only if a certain quadratic form of the normal observations is bounded by 0, and study the property of the corresponding coefficient matrix, particularly about its characteristic polynomial. In Section S1.2, we give an integral representation of the rejection probability. In Section S1.3, we study the first and second order partial derivatives of the rejection probability. In Section S1.4, we study bounds on the first order partial derivative of the rejection probability.

S1.1 Quadratic form and characteristic polynomial of the coefficient matrix

Lemma S1.1. Recall the definition of T_m in (8). Let $\tilde{\psi} \equiv (\psi_{m+1}, \psi_1, \dots, \psi_m)^\top$. Then for any $c > 0$,

$$|T_m| > c \iff \tilde{\psi}^\top \mathbf{V} \tilde{\psi} < 0 \quad (\text{S25})$$

where

$$\kappa \equiv \frac{mc^2}{m-1} \quad \text{and} \quad \mathbf{V} \equiv \begin{pmatrix} -m & \mathbf{1}_m^\top \\ \mathbf{1}_m & \kappa \mathbf{I}_m - \frac{\kappa+1}{m} \mathbf{1}_m \mathbf{1}_m^\top \end{pmatrix}.$$

Proof of Lemma S1.1. By definition, for any $c > 0$, $|T_m| > c$ if and only if $|\psi_{m+1} - \bar{\psi}_m| > cS_m$, which is further equivalent to $mc^2S_m^2 - m(\psi_{m+1} - \bar{\psi}_m)^2 < 0$. Let $\mathbf{1}_m$ be an m -dimensional vector with all elements being 1, $\mathbf{0}_m$ be an m -dimensional vector with all elements being 0, and \mathbf{I}_m be an $m \times m$ identity matrix. By some algebra, we can show that

$$(\psi_{m+1} - \bar{\psi}_m)^2 = \tilde{\psi}^\top \begin{pmatrix} 1 \\ -m^{-1} \mathbf{1}_m \end{pmatrix} \begin{pmatrix} 1 & -m^{-1} \mathbf{1}_m^\top \end{pmatrix} \tilde{\psi} = \tilde{\psi}^\top \begin{pmatrix} 1 & -m^{-1} \mathbf{1}_m^\top \\ -m^{-1} \mathbf{1}_m & m^{-2} \mathbf{1}_m \mathbf{1}_m^\top \end{pmatrix} \tilde{\psi},$$

and

$$\begin{aligned}
(m-1)S_m^2 &= \sum_{j=1}^m (\psi_j - \bar{\psi}_m)^2 = \boldsymbol{\psi}_{1:m}^\top (\mathbf{I}_m - m^{-1} \mathbf{1}_m \mathbf{1}_m^\top) \boldsymbol{\psi}_{1:m} = \boldsymbol{\psi}_{1:m}^\top (\mathbf{I}_m - m^{-1} \mathbf{1}_m \mathbf{1}_m^\top) \boldsymbol{\psi}_{1:m} \\
&= \tilde{\boldsymbol{\psi}}^\top \begin{pmatrix} 0 & \mathbf{0}_m^\top \\ \mathbf{0}_m & \mathbf{I}_m - m^{-1} \mathbf{1}_m \mathbf{1}_m^\top \end{pmatrix} \tilde{\boldsymbol{\psi}},
\end{aligned}$$

where $\boldsymbol{\psi}_{1:m} \equiv (\psi_1, \dots, \psi_m)^\top$. These imply that

$$\begin{aligned}
&mc^2 S_m^2 - m(\psi_{m+1} - \bar{\psi}_m)^2 \\
&= \frac{mc^2}{m-1} \tilde{\boldsymbol{\psi}}^\top \begin{pmatrix} 0 & \mathbf{0}_m^\top \\ \mathbf{0}_m & \mathbf{I}_m - m^{-1} \mathbf{1}_m \mathbf{1}_m^\top \end{pmatrix} \tilde{\boldsymbol{\psi}} - m \tilde{\boldsymbol{\psi}}^\top \begin{pmatrix} 1 & -m^{-1} \mathbf{1}_m^\top \\ -m^{-1} \mathbf{1}_m & m^{-2} \mathbf{1}_m \mathbf{1}_m^\top \end{pmatrix} \tilde{\boldsymbol{\psi}} \\
&= \tilde{\boldsymbol{\psi}}^\top \left[\begin{pmatrix} 0 & \mathbf{0}_m^\top \\ \mathbf{0}_m & \frac{mc^2}{m-1} \mathbf{I}_m - \frac{c^2}{m-1} \mathbf{1}_m \mathbf{1}_m^\top \end{pmatrix} - \begin{pmatrix} m & -\mathbf{1}_m^\top \\ -\mathbf{1}_m & m^{-1} \mathbf{1}_m \mathbf{1}_m^\top \end{pmatrix} \right] \tilde{\boldsymbol{\psi}} \\
&= \tilde{\boldsymbol{\psi}}^\top \begin{pmatrix} -m & \mathbf{1}_m^\top \\ \mathbf{1}_m & \frac{mc^2}{m-1} \mathbf{I}_m - \frac{1}{m} \left(\frac{mc^2}{m-1} + 1 \right) \mathbf{1}_m \mathbf{1}_m^\top \end{pmatrix} \tilde{\boldsymbol{\psi}} \\
&= \tilde{\boldsymbol{\psi}}^\top \mathbf{V} \tilde{\boldsymbol{\psi}},
\end{aligned}$$

where the last equality holds by definition. From the above, we derive Lemma S1.1. \square

Lemma S1.2. Let $\mathbf{D} \equiv \text{diag}(\sigma_{m+1}, \sigma_1, \dots, \sigma_m)$, and define \mathbf{V} and κ as in Lemma S1.1. For any nonnegative $\sigma_1^2, \sigma_2^2, \dots, \sigma_{m+1}^2$,

$$f(\lambda) \equiv -(m\sigma_{m+1}^2 + \lambda) \prod_{i=1}^m (\kappa\sigma_i^2 - \lambda) + \left(\kappa\sigma_{m+1}^2 + \frac{\kappa+1}{m} \lambda \right) \cdot \sum_{i=1}^m \left[\sigma_i^2 \prod_{j \neq i, 1 \leq j \leq m} (\kappa\sigma_j^2 - \lambda) \right]$$

is a characteristic polynomial of $\mathbf{D}\mathbf{V}\mathbf{D}$.

Proof of Lemma S1.2. We first consider the case where $\sigma_1, \sigma_2, \dots, \sigma_{m+1}$ are all positive. Let $\text{diag}(\sigma_{1:m}^{-2})$ be an $m \times m$ diagonal matrix with diagonal elements $(\sigma_1^{-2}, \dots, \sigma_m^{-2})$. By definition, we have

$$\begin{aligned}
\mathbf{V} - \lambda \mathbf{D}^{-2} &= \begin{pmatrix} -m & \mathbf{1}_m^\top \\ \mathbf{1}_m & \kappa \mathbf{I}_m - \frac{\kappa+1}{m} \mathbf{1}_m \mathbf{1}_m^\top \end{pmatrix} - \lambda \mathbf{D}^{-2} \\
&= \begin{pmatrix} -m - \lambda \sigma_{m+1}^{-2} & \mathbf{1}_m^\top \\ \mathbf{1}_m & \kappa \mathbf{I}_m - \lambda \text{diag}(\sigma_{1:m}^{-2}) - \frac{\kappa+1}{m} \mathbf{1}_m \mathbf{1}_m^\top \end{pmatrix}
\end{aligned}$$

$$= \begin{pmatrix} b & 1 & 1 & \cdots & 1 \\ 1 & d + c_1 & d & \cdots & d \\ 1 & d & d + c_2 & \cdots & d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & d & d & \cdots & d + c_m \end{pmatrix},$$

where $b = -m - \lambda\sigma_{m+1}^{-2}$, $d = -\frac{\kappa+1}{m}$, and $c_i = \kappa - \lambda\sigma_i^{-2}$ for $1 \leq i \leq m$. Applying [Bakirov \(1998, Proposition 1\)](#), we have

$$\begin{aligned} & \det(\mathbf{D}\mathbf{V}\mathbf{D} - \lambda\mathbf{I}_{m+1}) \\ &= \det(\mathbf{D})^2 \det(\mathbf{V} - \lambda\mathbf{D}^{-2}) \\ &= \prod_{j=1}^{m+1} \sigma_j^2 \begin{vmatrix} b & 1 & 1 & \cdots & 1 \\ 1 & d + c_1 & d & \cdots & d \\ 1 & d & d + c_2 & \cdots & d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & d & d & \cdots & d + c_m \end{vmatrix} \\ &= \prod_{j=1}^{m+1} \sigma_j^2 \cdot b \prod_{i=1}^m c_i \left(1 + d \sum_{i=1}^m \frac{1}{c_i} - \frac{1}{b} \sum_{i=1}^m \frac{1}{c_i} \right) \\ &= \sigma_{m+1}^2 b \cdot \prod_{i=1}^m (\sigma_i^2 c_i) \cdot \left[1 + \left(d - \frac{\sigma_{m+1}^2}{\sigma_{m+1}^2 b} \right) \sum_{i=1}^m \frac{\sigma_i^2}{c_i \sigma_i^2} \right] \\ &= -(m\sigma_{m+1}^2 + \lambda) \cdot \prod_{i=1}^m (\kappa\sigma_i^2 - \lambda) \cdot \left[1 + \left(-\frac{\kappa+1}{m} + \frac{\sigma_{m+1}^2}{m\sigma_{m+1}^2 + \lambda} \right) \sum_{i=1}^m \frac{\sigma_i^2}{\kappa\sigma_i^2 - \lambda} \right] \\ &= -(m\sigma_{m+1}^2 + \lambda) \prod_{i=1}^m (\kappa\sigma_i^2 - \lambda) - \left[-\frac{\kappa+1}{m} (m\sigma_{m+1}^2 + \lambda) + \sigma_{m+1}^2 \right] \cdot \sum_{i=1}^m \left[\sigma_i^2 \prod_{j \neq i, 1 \leq j \leq m} (\kappa\sigma_j^2 - \lambda) \right] \\ &= -(m\sigma_{m+1}^2 + \lambda) \prod_{i=1}^m (\kappa\sigma_i^2 - \lambda) + \left(\kappa\sigma_{m+1}^2 + \frac{\kappa+1}{m} \lambda \right) \cdot \sum_{i=1}^m \left[\sigma_i^2 \prod_{j \neq i, 1 \leq j \leq m} (\kappa\sigma_j^2 - \lambda) \right] \\ &= f(\lambda), \end{aligned}$$

where the last equality holds by definition.

We then consider the case where some of the $\{\sigma_j\}_{j=1}^{m+1}$ can be zero. By definition, for any $\lambda \in \mathbb{R}$, both $\det(\mathbf{D}\mathbf{V}\mathbf{D} - \lambda\mathbf{I}_{m+1})$ and $f(\lambda)$, viewed as functions of $\{\sigma_j\}_{j=1}^{m+1}$, are polynomial and thus continuous functions of $\{\sigma_j\}_{j=1}^{m+1}$. By continuity and the fact that $\det(\mathbf{D}\mathbf{V}\mathbf{D} - \lambda\mathbf{I}_{m+1}) = f(\lambda)$ for any positive $\{\sigma_j\}_{j=1}^{m+1}$, we can know that $\det(\mathbf{D}\mathbf{V}\mathbf{D} - \lambda\mathbf{I}_{m+1})$ must equal $f(\lambda)$ for any nonnegative $\{\sigma_j\}_{j=1}^{m+1}$.

From the above, Lemma S1.2 holds. \square

Lemma S1.3. Consider any positive σ_{m+1} and any nonnegative $\sigma_1, \dots, \sigma_m$. Let $\kappa \equiv \frac{mc^2}{m-1}$, $\gamma_i \equiv \frac{\sigma_i}{\sigma_{m+1}}$ for $1 \leq i \leq m$, and

$$g(\theta) = -(m + \theta) \prod_{i=1}^m (\kappa \gamma_i^2 - \theta) + \left(\kappa + \frac{\kappa + 1}{m} \theta \right) \cdot \sum_{i=1}^m \left[\gamma_i^2 \prod_{j \neq i} (\kappa \gamma_j^2 - \theta) \right]. \quad (\text{S26})$$

Then, for any $\lambda \in \mathbb{R}$, $f(\lambda)$ in Lemma S1.2 can be written equivalently as

$$f(\lambda) = (\sigma_{m+1}^2)^{m+1} g\left(\frac{\lambda}{\sigma_{m+1}^2}\right).$$

Consequently, if $\lambda_1, \dots, \lambda_{m+1}$ are the $m + 1$ roots of $f(\cdot)$, then $\theta_i = \frac{\lambda_i}{\sigma_{m+1}^2}$, $i = 1, 2, \dots, m + 1$, are the $(m + 1)$ roots of $g(\cdot)$.

Proof of Lemma S1.3. For any positive σ_{m+1}^2 , $\lambda \in \mathbb{R}$ and $\theta \equiv \frac{\lambda}{\sigma_{m+1}^2}$, by definition and some algebra, we can verify that

$$\begin{aligned} (\sigma_{m+1}^2)^{-(m+1)} \cdot f(\lambda) &= -(m + \theta) \prod_{i=1}^m (\kappa \gamma_i^2 - \theta) + \left(\kappa + \frac{\kappa + 1}{m} \theta \right) \cdot \sum_{i=1}^m \left[\gamma_i^2 \prod_{j \neq i} (\kappa \gamma_j^2 - \theta) \right] \\ &= g(\theta). \end{aligned}$$

This immediately implies Lemma S1.3. \square

Lemma S1.4. Consider any $c > 0$, any positive σ_{m+1} , and any nonnegative $\sigma_1, \dots, \sigma_m$. Define $\kappa, \gamma_1, \dots, \gamma_m$ and $g(\cdot)$ as in Lemma S1.3. Define further $\tau \equiv \frac{\kappa+1}{\kappa m}$. Let $\gamma_{(1)} \leq \gamma_{(2)} \leq \dots \leq \gamma_{(m)}$ be the sorted values of $\gamma_1, \gamma_2, \dots, \gamma_m$. The $(m + 1)$ roots of $g(\theta)$ in (S26) can then be characterized as follows:

- (a) for each $1 \leq i \leq m - 1$, there is a nonnegative root $\theta_i \in (\kappa \gamma_{(i)}^2, \kappa \gamma_{(i+1)}^2)$ if $\gamma_{(i)} < \gamma_{(i+1)}$ and $\theta_i = \kappa \gamma_{(i)}^2$ otherwise;
- (b) there is a zero root $\theta_m = 0$;
- (c) there is a negative root $\theta_{m+1} < -\tau^{-1}$.

Proof of Lemma S1.4. First, we give an equivalent expression for $g(\cdot)$. Let $\{\tilde{\gamma}_i\}_{i=1}^L$ be the unique ordered values of $\{\gamma_i\}_{i=1}^m$ such that $\tilde{\gamma}_1 < \dots < \tilde{\gamma}_L$. In addition, let m_i be the multiplicity of $\tilde{\gamma}_i$ for each $i = 1, \dots, L$. Note that $\sum_{i=1}^L m_i = m$ by construction. From

Lemma S1.3, $g(\cdot)$ in (S26) has the following equivalent forms:

$$\begin{aligned}
g(\theta) &= -(m + \theta) \prod_{i=1}^m (\kappa \gamma_i^2 - \theta) + \left(\kappa + \frac{\kappa + 1}{m} \theta \right) \cdot \sum_{i=1}^m \left[\gamma_i^2 \prod_{j \neq i} (\kappa \gamma_j^2 - \theta) \right] \\
&= \prod_{i=1}^L (\kappa \tilde{\gamma}_i^2 - \theta)^{m_i - 1} \left\{ -(m + \theta) \prod_{i=1}^L (\kappa \tilde{\gamma}_i^2 - \theta) + \left(\kappa + \frac{\kappa + 1}{m} \theta \right) \cdot \sum_{i=1}^L \left[m_i \tilde{\gamma}_i^2 \prod_{j \neq i} (\kappa \tilde{\gamma}_j^2 - \theta) \right] \right\} \\
&= \prod_{i=1}^L (\kappa \tilde{\gamma}_i^2 - \theta)^{m_i - 1} \tilde{g}(\theta), \tag{S27}
\end{aligned}$$

where

$$\tilde{g}(\theta) \equiv -(m + \theta) \prod_{i=1}^L (\kappa \tilde{\gamma}_i^2 - \theta) + \left(\kappa + \frac{\kappa + 1}{m} \theta \right) \cdot \sum_{i=1}^L \left[m_i \tilde{\gamma}_i^2 \prod_{j \neq i} (\kappa \tilde{\gamma}_j^2 - \theta) \right].$$

From (S27), it can be seen that there are $\sum_{i=1}^L (m_i - 1) = \sum_{i=1}^L m_i - L = m - L$ roots from the factor $\prod_{i=1}^L (\kappa \tilde{\gamma}_i^2 - \theta)^{m_i - 1}$.

Second, we prove (a) and (b) in the case where $\tilde{\gamma}_1 > 0$. Consider the value of $g(\cdot)$ evaluated at $\kappa \tilde{\gamma}_l^2$ for $l = 1, \dots, L$. By definition, for $1 \leq l \leq L$, we have

$$\begin{aligned}
\tilde{g}(\kappa \tilde{\gamma}_l^2) &\equiv \underbrace{-(m + \kappa \tilde{\gamma}_l^2) \prod_{i=1}^L (\kappa \tilde{\gamma}_i^2 - \kappa \tilde{\gamma}_l^2)}_{=0} + \left(\kappa + \frac{\kappa + 1}{m} \kappa \tilde{\gamma}_l^2 \right) \cdot \sum_{i=1}^L \left[m_i \tilde{\gamma}_i^2 \prod_{j \neq i} (\kappa \tilde{\gamma}_j^2 - \kappa \tilde{\gamma}_l^2) \right] \\
&= \left(\kappa + \frac{\kappa + 1}{m} \kappa \tilde{\gamma}_l^2 \right) \cdot m_l \tilde{\gamma}_l^2 \cdot \prod_{j \neq l} (\kappa \tilde{\gamma}_j^2 - \kappa \tilde{\gamma}_l^2) \\
&= \underbrace{\kappa^{L-1} \left(\kappa + \frac{\kappa + 1}{m} \kappa \tilde{\gamma}_l^2 \right) \cdot m_l \tilde{\gamma}_l^2}_{>0} \cdot \prod_{j \neq l} (\tilde{\gamma}_j^2 - \tilde{\gamma}_l^2). \tag{S28}
\end{aligned}$$

By the construction of $\{\tilde{\gamma}_j\}_{j=1}^L$, we can know that the sign of $\tilde{g}(\kappa \tilde{\gamma}_l^2)$ is $(-1)^{l-1}$ for $1 \leq l \leq L$. Thus, by the continuity of $\tilde{g}(\cdot)$, it must have a positive root in $(\kappa \tilde{\gamma}_l^2, \kappa \tilde{\gamma}_{l+1}^2)$, for $l = 1, 2, \dots, L - 1$. In addition, we can verify that

$$\begin{aligned}
\tilde{g}(0) &\equiv -m \prod_{i=1}^L (\kappa \tilde{\gamma}_i^2) + \kappa \cdot \sum_{i=1}^L \left[m_i \tilde{\gamma}_i^2 \prod_{j \neq i, 1 \leq j \leq L} (\kappa \tilde{\gamma}_j^2) \right] \\
&= -m \kappa^L \prod_{i=1}^L \tilde{\gamma}_i^2 + \kappa^L \cdot \sum_{i=1}^L \left(m_i \prod_{j=1}^L \tilde{\gamma}_j^2 \right)
\end{aligned}$$

$$= \kappa^L \prod_{i=1}^L \tilde{\gamma}_i^2 \cdot \left(-m + \sum_{i=1}^L m_i \right) = 0.$$

Combined with the first part, we can know that (a) and (b) in Lemma S1.4 hold.

Third, we prove (a) and (b) in the case where $\tilde{\gamma}_1 = 0$. In this case, $\tilde{g}(\cdot)$ simplifies to

$$\begin{aligned} \tilde{g}(\theta) &\equiv -(m + \theta)(-\theta) \prod_{i=2}^L (\kappa \tilde{\gamma}_i^2 - \theta) + \left(\kappa + \frac{\kappa + 1}{m} \theta \right) (-\theta) \cdot \sum_{i=2}^L \left[m_i \tilde{\gamma}_i^2 \prod_{j \neq 1, i} (\kappa \tilde{\gamma}_j^2 - \theta) \right] \\ &= -\theta \cdot \left\{ -(m + \theta) \prod_{i=2}^L (\kappa \tilde{\gamma}_i^2 - \theta) + \left(\kappa + \frac{\kappa + 1}{m} \theta \right) \cdot \sum_{i=2}^L \left[m_i \tilde{\gamma}_i^2 \prod_{j \neq 1, i} (\kappa \tilde{\gamma}_j^2 - \theta) \right] \right\} \\ &= -\theta \cdot \check{g}(\theta), \end{aligned} \tag{S29}$$

where

$$\check{g}(\theta) = -(m + \theta) \prod_{i=2}^L (\kappa \tilde{\gamma}_i^2 - \theta) + \left(\kappa + \frac{\kappa + 1}{m} \theta \right) \cdot \sum_{i=2}^L \left[m_i \tilde{\gamma}_i^2 \prod_{j \neq 1, i} (\kappa \tilde{\gamma}_j^2 - \theta) \right].$$

We can verify that

$$\begin{aligned} \check{g}(\kappa \tilde{\gamma}_1^2) &= \check{g}(0) = -m \prod_{i=2}^L (\kappa \tilde{\gamma}_i^2) + \kappa \cdot \sum_{i=2}^L \left[m_i \tilde{\gamma}_i^2 \prod_{j \neq 1, i} (\kappa \tilde{\gamma}_j^2) \right] = \kappa^{L-1} \prod_{i=2}^L \tilde{\gamma}_i^2 \left(-m + \sum_{i=2}^L m_i \right) \\ &= \kappa^{L-1} \prod_{i=2}^L \tilde{\gamma}_i^2 \cdot (-m_1) < 0, \end{aligned}$$

where the last equality holds because $\sum_{i=1}^L m_i = m$, and for $2 \leq l \leq L$,

$$\begin{aligned} \check{g}(\kappa \tilde{\gamma}_l^2) &= -(m + \kappa \tilde{\gamma}_l^2) \underbrace{\prod_{i=2}^L (\kappa \tilde{\gamma}_i^2 - \kappa \tilde{\gamma}_l^2)}_{=0} + \left(\kappa + \frac{\kappa + 1}{m} \kappa \tilde{\gamma}_l^2 \right) \cdot \sum_{i=2}^L \left[m_i \tilde{\gamma}_i^2 \prod_{j \neq 1, i} (\kappa \tilde{\gamma}_j^2 - \kappa \tilde{\gamma}_l^2) \right] \\ &= \underbrace{\left(\kappa + \frac{\kappa + 1}{m} \kappa \tilde{\gamma}_l^2 \right) \cdot m_l \tilde{\gamma}_l^2}_{>0} \cdot \prod_{j \neq 1, l} (\kappa \tilde{\gamma}_j^2 - \kappa \tilde{\gamma}_l^2), \end{aligned}$$

whose sign is $(-1)^{l-2}$. Thus, by the continuity of $\check{g}(\cdot)$, it must have a positive root in $(\kappa \tilde{\gamma}_l^2, \kappa \tilde{\gamma}_{l+1}^2)$, for $l = 1, 2, \dots, L-1$. In addition, as shown in (S29), $\tilde{g}(\cdot)$ also has a zero root. Combined with the first part, we can know that (a) and (b) in Lemma S1.4 holds.

Fourth, we prove (c). Consider first the sign of the limit of $g(\theta)$ as $\theta \rightarrow -\infty$. By

definition,

$$\begin{aligned}
& \lim_{\theta \rightarrow -\infty} \theta^{-(m+1)} g(\theta) \\
&= \lim_{\theta \rightarrow -\infty} \left\{ -(m/\theta + 1) \prod_{i=1}^m (\kappa \gamma_i^2 / \theta - 1) + \theta^{-1} \left(\kappa / \theta + \frac{\kappa + 1}{m} \right) \cdot \sum_{i=1}^m \left[\gamma_i^2 \prod_{j \neq i} (\kappa \gamma_j^2 / \theta - 1) \right] \right\} \\
&= (-1)^{m+1}.
\end{aligned}$$

Consequently, $g(\theta)$ is positive when θ is sufficiently small. Consider then the sign of $g(\theta)$ when $-\tau^{-1} \leq \theta < 0$. By definition, for $\theta < 0$,

$$\begin{aligned}
& \underbrace{\left[\prod_{i=1}^m (\kappa \gamma_i^2 - \theta) \right]^{-1}}_{>0} g(\theta) \\
&= -(m + \theta) + \left(\kappa + \frac{\kappa + 1}{m} \theta \right) \cdot \sum_{i=1}^m \frac{\gamma_i^2}{\kappa \gamma_i^2 - \theta} = -(m + \theta) + (1 + \tau \theta) \cdot \sum_{i=1}^m \frac{\kappa \gamma_i^2}{\kappa \gamma_i^2 - \theta} \\
&= \sum_{i=1}^m \left[\frac{\kappa \gamma_i^2 + \tau \theta \kappa \gamma_i^2}{\kappa \gamma_i^2 - \theta} - 1 \right] - \theta = \sum_{i=1}^m \frac{\tau \theta \kappa \gamma_i^2 + \theta}{\kappa \gamma_i^2 - \theta} - \theta = \theta \left(\sum_{i=1}^m \frac{\tau \kappa \gamma_i^2 + 1}{\kappa \gamma_i^2 - \theta} - 1 \right) \\
&= \theta \left[\sum_{i=1}^m \frac{\tau (\kappa \gamma_i^2 + \tau^{-1})}{\kappa \gamma_i^2 - \theta} - 1 \right]. \tag{S30}
\end{aligned}$$

When $-\tau^{-1} \leq \theta < 0$, we have $\tau^{-1} \geq -\theta$, and thus $\kappa \gamma_i^2 + \tau^{-1} \geq \kappa \gamma_i^2 - \theta$. This then implies that

$$\begin{aligned}
& \underbrace{\left[\prod_{i=1}^m (\kappa \gamma_i^2 - \theta) \right]^{-1}}_{>0} g(\theta) = \theta \left[\sum_{i=1}^m \frac{\tau (\kappa \gamma_i^2 + \tau^{-1})}{\kappa \gamma_i^2 - \theta} - 1 \right] \\
& \leq \theta \left(\sum_{i=1}^m \tau - 1 \right) \\
& = \theta (m\tau - 1) \\
& = \theta \left(\frac{\kappa + 1}{\kappa} - 1 \right) \\
& < 0.
\end{aligned}$$

Consequently, $g(\theta)$ is negative for $\theta \in [-\tau^{-1}, 0)$. From the above, $g(\cdot)$ must have a negative root in $(-\infty, -\tau^{-1})$. Thus, (c) in Lemma S1.4 holds.

From the above, Lemma S1.4 holds. \square

S1.2 Integral representation of the rejection probability

Lemma S1.5. Recall the test statistic T_m in (8). Let $\{\theta_i\}_{i=1}^{m+1}$ be the roots of the function $g(\cdot)$ in Lemma S1.3, with θ_{m+1} being the unique negative root, and $g_0(\theta) \equiv \prod_{i=1}^{m+1}(\theta - \theta_i)$. Then for any $c > 0$, any positive σ_{m+1} , and any nonnegative $\sigma_1, \dots, \sigma_m$,

$$\mathbb{P}[|T_m| > c] = \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{s^{\frac{m-1}{2}}}{[(-1)^m g_0(-s)]^{\frac{1}{2}}} ds = \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{s^{\frac{m-1}{2}}}{[-g(-s)]^{\frac{1}{2}}} ds$$

Proof of Lemma S1.5. Let $\{\tilde{\xi}_i\}_{i=1}^{m+1}$ be i.i.d. standard normal random variables, and $\boldsymbol{\xi} \equiv (\tilde{\xi}_1, \tilde{\xi}_2, \dots, \tilde{\xi}_{m+1})^\top$. From Lemma S1.1, the probability that $|T_m| > c$ is equivalent to:

$$\mathbb{P}[|T_m| > c] = \mathbb{P}[\tilde{\boldsymbol{\psi}}^\top \mathbf{V} \tilde{\boldsymbol{\psi}} < 0] = \mathbb{P}[\boldsymbol{\xi}^\top \mathbf{D} \mathbf{V} \mathbf{D} \boldsymbol{\xi} < 0] = \mathbb{P}\left[\sum_{i=1}^{m+1} \lambda_i \tilde{\xi}_i^2 < 0\right],$$

where $\lambda_1, \lambda_2, \dots, \lambda_{m+1}$ are eigenvalues of $\mathbf{D} \mathbf{V} \mathbf{D}$ or equivalently the roots of the characteristic polynomial in Lemma S1.2. From Lemma S1.3, it can be further written as

$$\mathbb{P}[|T_m| > c] = \mathbb{P}\left[\sum_{i=1}^{m+1} \lambda_i \tilde{\xi}_i^2 < 0\right] = \mathbb{P}\left[\sum_{i=1}^{m+1} \lambda_i \sigma_{m+1}^{-2} \tilde{\xi}_i^2 < 0\right] = \mathbb{P}\left[\sum_{i=1}^{m+1} \theta_i \tilde{\xi}_i^2 < 0\right].$$

Using Makshanov and Shalaevskii (1977, Theorem 2), we can write the probability as the following integral:

$$\mathbb{P}[|T_m| > c] = \mathbb{P}\left[\frac{\tilde{\xi}_{m+1}^2}{\sum_{i=1}^m \frac{\theta_i}{|\theta_{m+1}|} \tilde{\xi}_i^2} > 1\right] = \frac{1}{\pi} \int_0^\infty \frac{t^{-\frac{1}{2}} (1+t)^{-1}}{\sqrt{\prod_{i=1}^m [1 + \frac{\theta_i}{|\theta_{m+1}|} (1+t)]}} dt. \quad (\text{S31})$$

By some algebra and the definition of $g_0(\cdot)$,

$$\begin{aligned} & \prod_{i=1}^m \left[1 + \frac{\theta_i}{|\theta_{m+1}|} (1+t)\right] \\ &= \prod_{i=1}^m \left[\left(\frac{|\theta_{m+1}|}{1+t} + \theta_i\right) \frac{1+t}{|\theta_{m+1}|}\right] = \left(\frac{1+t}{|\theta_{m+1}|}\right)^m \prod_{i=1}^m \left(\frac{|\theta_{m+1}|}{1+t} + \theta_i\right) \\ &= \left(\frac{1+t}{|\theta_{m+1}|}\right)^m \left(\frac{|\theta_{m+1}|}{1+t} - |\theta_{m+1}|\right)^{-1} \prod_{i=1}^{m+1} \left(\frac{|\theta_{m+1}|}{1+t} + \theta_i\right) \end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{t} \left(\frac{1+t}{|\theta_{m+1}|} \right)^{m+1} \prod_{i=1}^{m+1} \left(\frac{|\theta_{m+1}|}{1+t} + \theta_i \right) = (-1)^{m+2} \frac{1}{t} \left(\frac{1+t}{|\theta_{m+1}|} \right)^{m+1} \prod_{i=1}^{m+1} \left(-\frac{|\theta_{m+1}|}{1+t} - \theta_i \right) \\
&= (-1)^m \frac{1}{t} \left(\frac{1+t}{|\theta_{m+1}|} \right)^{m+1} g_0 \left(-\frac{|\theta_{m+1}|}{1+t} \right).
\end{aligned}$$

We can then write (S31) as

$$\mathbb{P}[|T_m| > c] = \frac{1}{\pi} \int_0^\infty \frac{(1+t)^{-\frac{m+1}{2}-1} |\theta_{m+1}|^{\frac{m+1}{2}}}{[(-1)^m g_0(-\frac{|\theta_{m+1}|}{1+t})]^{\frac{1}{2}}} dt.$$

After a change of variable with $s = \frac{|\theta_{m+1}|}{1+t}$ and $ds = -|\theta_{m+1}|(1+t)^{-2} dt$, this further simplifies to

$$\mathbb{P}[|T_m| > c] = \frac{1}{\pi} \int_{|\theta_{m+1}|}^0 \frac{(1+t)^{-1} s^{\frac{m+1}{2}}}{[(-1)^m g_0(-s)]^{\frac{1}{2}}} \cdot -\frac{(1+t)^2}{|\theta_{m+1}|} ds = \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{s^{\frac{m-1}{2}}}{[(-1)^m g_0(-s)]^{\frac{1}{2}}} ds.$$

Note that, by definition, $g(\theta)$ and $g_0(\theta)$ are both polynomial functions of θ and have exactly the same roots. Hence, they must be a scalar multiple of each other. Because the coefficient of θ^{m+1} in $g(\theta)$ is $(-1)^{m+1}$ and the coefficient of θ^{m+1} in $g_0(\theta)$ is 1, we have $g(\theta) = (-1)^{m+1} \cdot g_0(\theta)$. Consequently,

$$\mathbb{P}[|T_m| > c] = \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{s^{\frac{m-1}{2}}}{[(-1)^m g_0(-s)]^{\frac{1}{2}}} ds = \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{s^{\frac{m-1}{2}}}{[-g(-s)]^{\frac{1}{2}}} ds$$

From the above, Lemma S1.5 holds. □

Lemma S1.6. Recall the test statistic T_m in (8), and consider any $c > 0$, any positive σ_{m+1} , and any nonnegative $\sigma_1, \dots, \sigma_m$. Define $\kappa \equiv \frac{mc^2}{m-1}$, $\tau \equiv \frac{\kappa+1}{\kappa m}$, and $\gamma_i \equiv \frac{\sigma_i}{\sigma_{m+1}}$ for $i \in \{1, \dots, m\}$. Let θ_{m+1} be the unique negative root of the function $g(\cdot)$ in Lemma S1.3. Let $x_i \equiv \kappa \gamma_i^2$ for $i \in \{1, \dots, m\}$. Then,

$$\mathbb{P}[|T_m| > c] = \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{U(\mathbf{x}, s)}{\sqrt{|\theta_{m+1}| - s}} ds,$$

where $\mathbf{x} \equiv (x_1, \dots, x_m)$,

$$U(\mathbf{x}, s) \equiv \frac{s^{\frac{m}{2}-1}}{\sqrt{P(\mathbf{x}, s)Q(\mathbf{x}, s)}}, \quad P(\mathbf{x}, s) \equiv \sum_{i=1}^m \frac{1 + \tau x_i}{(x_i + s)(x_i - \theta_{m+1})}, \quad Q(\mathbf{x}, s) \equiv \prod_{i=1}^m (x_i + s).$$

Proof of Lemma S1.6. By the same logic as (S30), for $\theta < 0$, we have

$$g(\theta) = \prod_{i=1}^m (\kappa \gamma_i^2 - \theta) \cdot \theta \left(\sum_{i=1}^m \frac{\tau \kappa \gamma_i^2 + 1}{\kappa \gamma_i^2 - \theta} - 1 \right) = \prod_{i=1}^m (x_i - \theta) \cdot \theta \left(\sum_{i=1}^m \frac{\tau x_i + 1}{x_i - \theta} - 1 \right). \quad (\text{S32})$$

Because $g(\theta_{m+1}) = 0$ and $\theta_{m+1} < 0$, we must have

$$1 = \sum_{i=1}^m \frac{1 + \tau x_i}{x_i - \theta_{m+1}}. \quad (\text{S33})$$

Substituting (S33) into (S32), we can write $g(\theta)$ for $\theta < 0$ as

$$\begin{aligned} g(\theta) &= \prod_{i=1}^m (x_i - \theta) \cdot \theta \left(\sum_{i=1}^m \frac{\tau x_i + 1}{x_i - \theta} - \sum_{i=1}^m \frac{1 + \tau x_i}{x_i - \theta_{m+1}} \right) \\ &= \prod_{i=1}^m (x_i - \theta) \cdot \theta \left[\sum_{i=1}^m (1 + \tau x_i) \left(\frac{1}{x_i - \theta} - \frac{1}{x_i - \theta_{m+1}} \right) \right] \\ &= \prod_{i=1}^m (x_i - \theta) \cdot \theta (\theta - \theta_{m+1}) \sum_{i=1}^m \frac{1 + \tau x_i}{(x_i - \theta)(x_i - \theta_{m+1})} \end{aligned} \quad (\text{S34})$$

Substituting (S34) into the integral representation in Lemma S1.5, we then have

$$\begin{aligned} \mathbb{P}[|T_m| > c] &= \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{s^{\frac{m-1}{2}}}{\sqrt{-\prod_{i=1}^m (x_i + s)(-s)(-s - \theta_{m+1}) \sum_{i=1}^m \frac{1 + \tau x_i}{(x_i + s)(x_i - \theta_{m+1})}}} ds \\ &= \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{s^{\frac{m}{2}-1}}{\sqrt{\prod_{i=1}^m (x_i + s)(-s - \theta_{m+1}) \sum_{i=1}^m \frac{1 + \tau x_i}{(x_i + s)(x_i - \theta_{m+1})}}} ds \\ &= \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{s^{\frac{m}{2}-1}}{\sqrt{Q(\mathbf{x}, s)P(\mathbf{x}, s)(|\theta_{m+1}| - s)}} ds \\ &= \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{U(\mathbf{x}, s)}{\sqrt{|\theta_{m+1}| - s}} ds. \end{aligned}$$

Therefore, Lemma S1.6 holds. □

S1.3 Derivatives of the rejection probability

Lemma S1.7. Consider the same setup and notations in Lemma S1.6. Fix c, x_3, \dots, x_m and θ_{m+1} , and let $z = x_1 > 0$ and $w = x_2 > 0$, where we view w as a function of z as implied by

(S33). Define $\Delta \equiv \frac{z-w}{(z-\theta_{m+1})^2}$, $A \equiv wz - \theta_{m+1}^2$ and $B \equiv z + w - 2\theta_{m+1}$. We then have

$$\frac{\partial w}{\partial z} = -\frac{(w - \theta_{m+1})^2}{(z - \theta_{m+1})^2}, \quad (\text{S35})$$

$$\frac{\partial P(\mathbf{x}, s)}{\partial z} = \frac{2\Delta(1 - \tau s)[A + \frac{B}{2}(s - \theta_{m+1})]}{(z + s)^2(w + s)^2}, \quad (\text{S36})$$

$$\frac{\partial Q(\mathbf{x}, s)}{\partial z} = \Delta(A + Bs) \prod_{i=3}^m (x_i + s). \quad (\text{S37})$$

Proof of Lemma S1.7. From Lemma S1.4, $\theta_{m+1} < -\tau^{-1}$, which implies that $-\tau\theta_{m+1} > 1$. Consequently, $\frac{1+\tau w}{w-\theta_{m+1}}$ is increasing in w , because

$$\frac{\partial}{\partial w} \left(\frac{1 + \tau w}{w - \theta_{m+1}} \right) = \frac{-\tau\theta_{m+1} - 1}{(w - \theta_{m+1})^2} > 0.$$

This ensures that w can be uniquely determined by z through (S33), once we fix c, x_3, \dots, x_m and θ_{m+1} . Below we prove the three derivatives in (S35)–(S37) separately as follows.

- **Proof of (S35).** From (S33), we have

$$1 = \frac{1 + \tau z}{z - \theta_{m+1}} + \frac{1 + \tau w}{w - \theta_{m+1}} + \underbrace{\sum_{i=3}^m \frac{1 + \tau x_i}{x_i - \theta_{m+1}}}_{\text{"Constant"}}.$$

Taking derivative of the above gives

$$0 = \frac{-\tau\theta_{m+1} - 1}{(z - \theta_{m+1})^2} dz + \frac{-\tau\theta_{m+1} - 1}{(w - \theta_{m+1})^2} dw.$$

As discussed before, $-\tau\theta_{m+1} > 1$. We then have

$$\frac{\partial w}{\partial z} = -\frac{(w - \theta_{m+1})^2}{(z - \theta_{m+1})^2},$$

i.e., (S35) holds.

- **Proof of (S36).** Let $\tilde{p}(x) \equiv \frac{1+\tau x}{(x+s)(x-\theta_{m+1})}$. We have

$$\frac{d\tilde{p}(x)}{dx} = -\frac{\tau x^2 + 2x + (s - \theta_{m+1} + s\theta_{m+1}\tau)}{(x+s)^2(x-\theta_{m+1})^2} = -\frac{\tau x^2 + 2x + d}{(x+s)^2(x-\theta_{m+1})^2},$$

where $d \equiv s - \theta_{m+1} + s\theta_{m+1}\tau = s(1 + \theta_{m+1}\tau) - \theta_{m+1}$. Using (S35), we have

$$\begin{aligned} \frac{\partial P(\mathbf{x}, s)}{\partial z} &= \frac{d\tilde{p}(z)}{dz} + \frac{d\tilde{p}(w)}{dw} \frac{\partial w}{\partial z} \\ &= -\frac{\tau z^2 + 2z + d}{(z+s)^2(z-\theta_{m+1})^2} + \frac{\tau w^2 + 2w + d}{(w+s)^2(w-\theta_{m+1})^2} \frac{(w-\theta_{m+1})^2}{(z-\theta_{m+1})^2} \\ &= \frac{-(\tau z^2 + 2z + d)(w+s)^2 + (\tau w^2 + 2w + d)(z+s)^2}{(w+s)^2(z+s)^2(z-\theta_{m+1})^2}. \end{aligned} \quad (\text{S38})$$

The numerator of (S38) can be simplified as follows:

$$\begin{aligned} &-(\tau z^2 + 2z + d)(w+s)^2 + (\tau w^2 + 2w + d)(z+s)^2 \\ &= (z-w)[2wz(1-\tau s) - (w+z)(\tau s^2 - d) + 2ds - 2s^2] \\ &= (z-w)\{2wz(1-\tau s) - (w+z)[\tau s^2 - s(1+\theta_{m+1}\tau) + \theta_{m+1}] \\ &\quad + 2s^2(1+\theta_{m+1}\tau) - 2s\theta_{m+1} - 2s^2\} \\ &= (z-w)\{\underbrace{(zw - \theta_{m+1}^2)}_{=A}[2(1-\tau s)] - \underbrace{(w+z-2\theta_{m+1})}_{=B}[\tau s^2 - s(1+\theta_{m+1}\tau) + \theta_{m+1}] \\ &\quad + \underbrace{2s^2(1+\theta_{m+1}\tau) - 2s\theta_{m+1} - 2s^2 + 2\theta_{m+1}^2(1-\tau s) - 2\theta_{m+1}[\tau s^2 - s(1+\theta_{m+1}\tau) + \theta_{m+1}]}_{=0}\} \\ &= (z-w)\{A[2(1-\tau s)] - B[\tau s^2 - s(1+\theta_{m+1}\tau) + \theta_{m+1}]\} \\ &= (z-w)\{A[2(1-\tau s)] - B(\tau s - 1)(s - \theta_{m+1})\} \\ &= 2(z-w)(1-\tau s) \left[A + \frac{B}{2}(s - \theta_{m+1}) \right], \end{aligned} \quad (\text{S39})$$

where the definitions of A and B follow from the statement of the lemma. Next, substitute (S39) into (S38) and using the definition of Δ , we have

$$\frac{\partial P(\mathbf{x}, s)}{\partial z} = \frac{2(z-w)(1-\tau s) \left[A + \frac{B}{2}(s - \theta_{m+1}) \right]}{(w+s)^2(z+s)^2(z-\theta_{m+1})^2} = 2\Delta(1-\tau s) \frac{[A + \frac{B}{2}(s - \theta_{m+1})]}{(w+s)^2(z+s)^2}.$$

Thus, (S36) holds.

- **Proof of (S37).** By the definition of Q , we have

$$Q(\mathbf{x}, s) = (w+s)(z+s) \underbrace{\prod_{i=3}^m (x_i + s)}_{\text{"Constant"}}.$$

Using (S35), we have

$$\begin{aligned}
\frac{\partial[(w+s)(z+s)]}{\partial z} &= (w+s) + (z+s) \frac{\partial w}{\partial z} = (w+s) - (z+s) \frac{(w-\theta_{m+1})^2}{(z-\theta_{m+1})^2} \\
&= \frac{(z-w)[(wz - \theta_{m+1}^2) + s(z+w-2\theta_{m+1})]}{(z-\theta_{m+1})^2} \\
&= \Delta(A + Bs).
\end{aligned} \tag{S40}$$

Thus,

$$\frac{\partial Q(\mathbf{x}, s)}{\partial z} = \Delta(A + Bs) \prod_{i=1}^m (x_i + s),$$

i.e., (S37) holds.

From the above, Lemma S1.7 holds. □

Lemma S1.8. *Consider the same setup and notations as in Lemmas S1.6 and S1.7. We have*

$$\frac{\partial U(\mathbf{x}, s)}{\partial z} = -\frac{\Delta \tilde{h}(\mathbf{x}, s) U(\mathbf{x}, s)}{2},$$

where

$$\begin{aligned}
\tilde{h}(\mathbf{x}, s) &\equiv \tilde{h}_P(\mathbf{x}, s) + \tilde{h}_Q(\mathbf{x}, s), \\
\tilde{h}_P(\mathbf{x}, s) &\equiv \frac{2(1-\tau s)[A + \frac{B}{2}(s - \theta_{m+1})]}{(z+s)^2(w+s)^2 P(\mathbf{x}, s)}, \\
\tilde{h}_Q(\mathbf{x}, s) &\equiv \frac{A + Bs}{(w+s)(z+s)}.
\end{aligned}$$

Proof of Lemma S1.8. By definition,

$$\begin{aligned}
\frac{\partial U(\mathbf{x}, s)}{\partial z} &= -\frac{s^{\frac{m}{2}-1}}{2[P(\mathbf{x}, s)Q(\mathbf{x}, s)]^{\frac{3}{2}}} \left[Q(\mathbf{x}, s) \frac{\partial P(\mathbf{x}, s)}{\partial z} + P(\mathbf{x}, s) \frac{\partial Q(\mathbf{x}, s)}{\partial z} \right] \\
&= -\frac{U(\mathbf{x}, s)}{2} \left[\underbrace{\frac{1}{P(\mathbf{x}, s)} \frac{\partial P(\mathbf{x}, s)}{\partial z}}_{\equiv h_P(\mathbf{x}, s)} + \underbrace{\frac{1}{Q(\mathbf{x}, s)} \frac{\partial Q(\mathbf{x}, s)}{\partial z}}_{\equiv h_Q(\mathbf{x}, s)} \right].
\end{aligned} \tag{S41}$$

From Lemma S1.7, $h_P(\mathbf{x}, s)$ and $h_Q(\mathbf{x}, s)$ have the following equivalent forms:

$$\begin{aligned} h_P(\mathbf{x}, s) &= \frac{2\Delta(1 - \tau s)[A + \frac{B}{2}(s - \theta_{m+1})]}{(z + s)^2(w + s)^2 P(\mathbf{x}, s)} = \Delta \tilde{h}_P(\mathbf{x}, s), \\ h_Q(\mathbf{x}, s) &= \frac{\Delta(A + Bs) \prod_{i=3}^m (x_i + s)}{Q(\mathbf{x}, s)} = \frac{\Delta(A + Bs)}{(w + s)(z + s)} = \Delta \tilde{h}_Q(\mathbf{x}, s), \end{aligned}$$

where the last equality in each of the above two equations follows by definition. Substituting the above into (S41), we then have

$$\begin{aligned} \frac{\partial U(\mathbf{x}, s)}{\partial z} &= -\frac{U(\mathbf{x}, s)}{2} [h_P(\mathbf{x}, s) + h_Q(\mathbf{x}, s)] \\ &= -\frac{U(\mathbf{x}, s)}{2} \Delta [\tilde{h}_P(\mathbf{x}, s) + \tilde{h}_Q(\mathbf{x}, s)] \\ &= -\frac{U(\mathbf{x}, s)}{2} \Delta \tilde{h}(\mathbf{x}, s), \end{aligned}$$

where the last equality holds by definition. Therefore, Lemma S1.8 holds. \square

Lemma S1.9. Consider the same setup and notations as in Lemmas S1.6 and S1.7. Let $\delta \equiv A + \frac{B}{2}(s - \theta_{m+1})$. Then, we have

$$\frac{\partial A}{\partial z} = \Delta A, \tag{S42}$$

$$\frac{\partial B}{\partial z} = \Delta B, \tag{S43}$$

$$\frac{\partial \delta}{\partial z} = \Delta \delta. \tag{S44}$$

Proof of Lemma S1.9. We prove the three derivatives in (S42)–(S44) separately as follows.

• **Proof of (S42):**

$$\frac{\partial A}{\partial z} = w + z \frac{\partial w}{\partial z} = w - z \frac{(w - \theta_{m+1})^2}{(z - \theta_{m+1})^2} = \Delta(wz - \theta_{m+1}^2) = \Delta A.$$

• **Proof of (S43):**

$$\frac{\partial B}{\partial z} = 1 + \frac{\partial w}{\partial z} = 1 - \frac{(w - \theta_{m+1})^2}{(z - \theta_{m+1})^2} = \Delta B.$$

- **Proof of (S44):**

$$\frac{\partial \delta}{\partial z} = \frac{\partial A}{\partial z} + \frac{\partial B}{\partial z} \frac{(s - \theta_{m+1})}{2} = \Delta A + \frac{\Delta B}{2}(s - \theta_{m+1}) = \Delta \delta.$$

From the above, Lemma S1.9 holds. \square

Lemma S1.10. *Consider the same setup and notations as in Lemmas S1.6–S1.9. We have*

$$\frac{\partial \tilde{h}_P(\mathbf{x}, s)}{\partial z} = \Delta \tilde{h}_P(\mathbf{x}, s)[1 - \tilde{h}_P(\mathbf{x}, s) - 2\tilde{h}_Q(\mathbf{x}, s)], \quad (\text{S45})$$

$$\frac{\partial \tilde{h}_Q(\mathbf{x}, s)}{\partial z} = \Delta[\tilde{h}_Q(\mathbf{x}, s) - \tilde{h}_Q(\mathbf{x}, s)^2]. \quad (\text{S46})$$

Proof of Lemma S1.10. We prove the two derivatives in (S45) and (S46) separately as follows.

- **Proof of (S45):** Let $\tilde{\delta} \equiv 2(1 - \tau s)\delta$. By definition,

$$\begin{aligned} \tilde{h}_P(\mathbf{x}, s) &= \frac{2(1 - \tau s)[A + \frac{B}{2}(s - \theta_{m+1})]}{(z + s)^2(w + s)^2P(\mathbf{x}, s)} \\ &= \frac{2(1 - \tau s)\delta}{(z + s)^2(w + s)^2P(\mathbf{x}, s)} \\ &= \frac{\tilde{\delta}}{(z + s)^2(w + s)^2P(\mathbf{x}, s)}. \end{aligned}$$

Thus,

$$\begin{aligned} \frac{\partial \tilde{h}_P(\mathbf{x}, s)}{\partial z} &= \frac{\partial}{\partial z} \frac{\tilde{\delta}}{(w + s)^2(z + s)^2P(\mathbf{x}, s)} \\ &= \frac{1}{(w + s)^4(z + s)^4P(\mathbf{x}, s)^2} \left\{ [(w + s)^2(z + s)^2P(\mathbf{x}, s)] \frac{\partial \tilde{\delta}}{\partial z} \right. \\ &\quad \left. - \tilde{\delta} \frac{\partial [(w + s)^2(z + s)^2P(\mathbf{x}, s)]}{\partial z} \right\}. \quad (\text{S47}) \end{aligned}$$

From Lemma S1.9, we have

$$\frac{\partial \tilde{\delta}}{\partial z} = 2(1 - \tau s) \frac{\partial \delta}{\partial z} = 2\Delta(1 - \tau s)\delta = \Delta \tilde{\delta}. \quad (\text{S48})$$

Next, using the definition of δ in Lemma S1.9 and using the expression of $\frac{\partial P(\mathbf{x}, s)}{\partial z}$

from Lemma S1.7, we obtain

$$\begin{aligned}
& \frac{\partial[(z+s)^2(w+s)^2P(\mathbf{x},s)]}{\partial z} \\
&= (z+s)^2(w+s)^2 \frac{\partial P(\mathbf{x},s)}{\partial z} + P(\mathbf{x},s) \frac{\partial[(z+s)^2(w+s)^2]}{\partial z} \\
&= (z+s)^2(w+s)^2 \frac{\partial P(\mathbf{x},s)}{\partial z} + 2P(\mathbf{x},s)(z+s)(w+s) \frac{\partial[(z+s)(w+s)]}{\partial z} \\
&= (z+s)^2(w+s)^2 \frac{\Delta\tilde{\delta}}{(z+s)^2(w+s)^2} + 2P(\mathbf{x},s)(z+s)(w+s)\Delta(A+B s) \\
&= \Delta[\tilde{\delta} + 2P(\mathbf{x},s)(z+s)(w+s)(A+B s)], \tag{S49}
\end{aligned}$$

where the second last equality follows from (S40). From (S47)–(S49), we then have

$$\begin{aligned}
& \frac{\partial\tilde{h}_P(\mathbf{x},s)}{\partial z} \\
&= \frac{\Delta\tilde{\delta}[(w+s)^2(z+s)^2P(\mathbf{x},s)] - \Delta\tilde{\delta}[\tilde{\delta} + 2P(\mathbf{x},s)(z+s)(w+s)(A+B s)]}{(w+s)^4(z+s)^4P(\mathbf{x},s)^2} \\
&= \frac{\Delta\tilde{\delta}[(w+s)^2(z+s)^2P(\mathbf{x},s) - \tilde{\delta} - 2P(\mathbf{x},s)(z+s)(w+s)(A+B s)]}{(w+s)^4(z+s)^4P(\mathbf{x},s)^2} \\
&= \frac{\Delta\tilde{\delta}}{(w+s)^2(z+s)^2P(\mathbf{x},s)} \left[1 - \frac{\tilde{\delta}}{(w+s)^2(z+s)^2P(\mathbf{x},s)} - \frac{2(A+B s)}{(w+s)(z+s)} \right] \\
&= \Delta\tilde{h}_P(\mathbf{x},s) [1 - \tilde{h}_P(\mathbf{x},s) - 2\tilde{h}_Q(\mathbf{x},s)],
\end{aligned}$$

where the first equality follows from (S47)–(S49), the second and third equality follows from rearranging and simplifying the fractions, and the last equality follows from the definition of $\tilde{h}_P(\mathbf{x},s)$ and $\tilde{h}_Q(\mathbf{x},s)$.

- **Proof of (S46):** From the definition of $\tilde{h}_Q(\mathbf{x},s)$ in Lemma S1.8, we have

$$\begin{aligned}
\frac{\partial\tilde{h}_Q(\mathbf{x},s)}{\partial z} &= \frac{\partial}{\partial z} \frac{(A+B s)}{(w+s)(z+s)} = \frac{(w+s)(z+s) \frac{\partial(A+B s)}{\partial z} - (A+B s) \frac{\partial[(w+s)(z+s)]}{\partial z}}{(w+s)^2(z+s)^2} \\
&= \frac{(w+s)(z+s)\Delta(A+B s) - (A+B s)\Delta(A+B s)}{(w+s)^2(z+s)^2} \\
&= \frac{\Delta(A+B s)}{(w+s)(z+s)} - \frac{\Delta(A+B s)^2}{(w+s)^2(z+s)^2} \\
&= \Delta[\tilde{h}_Q(\mathbf{x},s) - \tilde{h}_Q(\mathbf{x},s)^2],
\end{aligned}$$

where the third last equality follows from (S40), and the last equality follows from the definition of $\tilde{h}_Q(\mathbf{x},s)$.

From the above, Lemma S1.10 holds. \square

Lemma S1.11. *Consider the same setup and notations in Lemmas S1.6–S1.9. For any function $r(\mathbf{x}, s)$ of \mathbf{x} and s , define*

$$L(r(\mathbf{x}, s)) \equiv \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{r(\mathbf{x}, s) U(\mathbf{x}, s)}{\sqrt{|\theta_{m+1}| - s}} ds.$$

From Lemma S1.6, $\mathbb{P}[|T_m| > c] = L(1)$, where 1 here denotes a constant function taking value one.

(i) *The first and second order derivatives of $\mathbb{P}[|T_m| > c]$ over z have the following equivalent forms:*

$$\frac{\partial \mathbb{P}[|T_m| > c]}{\partial z} = -\frac{\Delta}{2} L(\tilde{h}(\mathbf{x}, s)), \quad (\text{S50})$$

$$\frac{\partial^2 \mathbb{P}[|T_m| > c]}{\partial z^2} = -\frac{L(\tilde{h}(\mathbf{x}, s))}{2} \frac{\partial \Delta}{\partial z} - \frac{\Delta^2}{2} L(\tilde{h}(\mathbf{x}, s)) + \frac{3\Delta^2}{4} L(\tilde{h}(\mathbf{x}, s)^2). \quad (\text{S51})$$

(ii) *If $c \neq m^{-1/2}$, $z \neq w$, and $\frac{\partial \mathbb{P}[|T_m| > c]}{\partial z} = 0$, then we must have $\frac{\partial^2 \mathbb{P}[|T_m| > c]}{\partial z^2} > 0$.*

Proof of Lemma S1.11. We first prove (i). We prove the two derivatives in (S50) and (S51) separately as follows.

- **Proof of (S50):** From Lemmas S1.6 and S1.8, the derivative of $\mathbb{P}[|T_m| > c]$ with respect to z is

$$\begin{aligned} \frac{\partial \mathbb{P}[|T_m| > c]}{\partial z} &= \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{\partial U(\mathbf{x}, s)}{\partial z} \frac{1}{\sqrt{|\theta_{m+1}| - s}} ds \\ &= -\frac{\Delta}{2\pi} \int_0^{|\theta_{m+1}|} \frac{\tilde{h}(\mathbf{x}, s) U(\mathbf{x}, s)}{\sqrt{|\theta_{m+1}| - s}} ds \\ &= -\frac{\Delta}{2} L(\tilde{h}(\mathbf{x}, s)). \end{aligned}$$

- **Proof of (S51):** Using (S50) and recognizing that Δ , \tilde{h} , and U are functions of z , we have

$$\begin{aligned} &\frac{\partial^2 \mathbb{P}[|T_m| > c]}{\partial z^2} \\ &= -\frac{L(\tilde{h}(\mathbf{x}, s))}{2} \frac{\partial \Delta}{\partial z} - \frac{\Delta}{2\pi} \int_0^{|\theta_{m+1}|} \left[\frac{\partial \tilde{h}(\mathbf{x}, s)}{\partial z} U(\mathbf{x}, s) + \frac{\partial U(\mathbf{x}, s)}{\partial z} \tilde{h}(\mathbf{x}, s) \right] \frac{1}{\sqrt{|\theta_{m+1}| - s}} ds \end{aligned}$$

$$\begin{aligned}
&= -\frac{L(\tilde{h}(\mathbf{x}, s))}{2} \frac{\partial \Delta}{\partial z} - \frac{\Delta}{2\pi} \int_0^{|\theta_{m+1}|} \frac{\partial \tilde{h}(\mathbf{x}, s)}{\partial z} \frac{U(\mathbf{x}, s)}{\sqrt{|\theta_{m+1}| - s}} ds \\
&\quad + \frac{\Delta^2}{4\pi} \int_0^{|\theta_{m+1}|} \tilde{h}(\mathbf{x}, s)^2 \frac{U(\mathbf{x}, s)}{\sqrt{|\theta_{m+1}| - s}} ds \\
&= -\frac{L(\tilde{h}(\mathbf{x}, s))}{2} \frac{\partial \Delta}{\partial z} - \frac{\Delta}{2} L \left(\frac{\partial \tilde{h}(\mathbf{x}, s)}{\partial z} \right) + \frac{\Delta^2}{4} L(\tilde{h}(\mathbf{x}, s)^2), \tag{S52}
\end{aligned}$$

where the second last equality follows from Lemma S1.8. By the definition of $\tilde{h}(\mathbf{x}, s)$ in Lemma S1.8 and from Lemma S1.10, we have

$$\begin{aligned}
\frac{\partial \tilde{h}(\mathbf{x}, s)}{\partial z} &= \frac{\partial \tilde{h}_P(\mathbf{x}, s)}{\partial z} + \frac{\partial \tilde{h}_Q(\mathbf{x}, s)}{\partial z} \\
&= \Delta \tilde{h}_P(\mathbf{x}, s) [1 - \tilde{h}_P(\mathbf{x}, s) - 2\tilde{h}_Q(\mathbf{x}, s)] + \Delta [\tilde{h}_Q(\mathbf{x}, s) - \tilde{h}_Q(\mathbf{x}, s)^2] \\
&= \Delta \left\{ [\tilde{h}_P(\mathbf{x}, s) + \tilde{h}_Q(\mathbf{x}, s)] - [\tilde{h}_P(\mathbf{x}, s) + \tilde{h}_Q(\mathbf{x}, s)]^2 \right\} \\
&= \Delta [\tilde{h}(\mathbf{x}, s) - \tilde{h}(\mathbf{x}, s)^2]. \tag{S53}
\end{aligned}$$

Substituting (S53) into (S52) gives

$$\frac{\partial^2 \mathbb{P}[|T_m| > c]}{\partial z^2} = -\frac{L(\tilde{h}(\mathbf{x}, s))}{2} \frac{\partial \Delta}{\partial z} - \frac{\Delta^2}{2} L(\tilde{h}(\mathbf{x}, s)) + \frac{3\Delta^2}{4} L(\tilde{h}(\mathbf{x}, s)^2).$$

We now prove (ii). Because $z \neq w$, Δ is nonzero. From (S50) and the condition in (ii), $L(\tilde{h}(\mathbf{x}, s))$ must be zero. From (S51), we then have

$$\frac{\partial^2 \mathbb{P}[|T_m| > c]}{\partial z^2} = -\frac{L(\tilde{h}(\mathbf{x}, s))}{2} \frac{\partial \Delta}{\partial z} - \frac{\Delta^2}{2} L(\tilde{h}(\mathbf{x}, s)) + \frac{3\Delta^2}{4} L(\tilde{h}(\mathbf{x}, s)^2) = \frac{3\Delta^2}{4} L(\tilde{h}(\mathbf{x}, s)^2) \geq 0.$$

From the definition of $U(\mathbf{x}, s)$ in Lemma S1.6, we can know that $U(\mathbf{x}, s)$ is positive for $s \in (0, |\theta_{m+1}|)$. Therefore, $L(\tilde{h}(\mathbf{x}, s)^2) = 0$ if and only if $\tilde{h}(\mathbf{x}, s) = 0$ for all $s \in (0, |\theta_{m+1}|)$. Note that, by the definition of $\tilde{h}(\mathbf{x}, s)$ in Lemma S1.8, for any $s \in (0, |\theta_{m+1}|)$,

$$\begin{aligned}
&\tilde{h}(\mathbf{x}, s) = 0 \\
&\iff \frac{2(1 - \tau s)[A + \frac{B}{2}(s - \theta_{m+1})]}{(z + s)^2(w + s)^2 P(\mathbf{x}, s)} + \frac{A + Bs}{(w + s)(z + s)} = 0 \\
&\iff 2(1 - \tau s)[A + \frac{B}{2}(s - \theta_{m+1})] + (A + Bs)(z + s)(w + s)P(\mathbf{x}, s) = 0 \\
&\iff \chi(s) = 0,
\end{aligned}$$

where

$$\begin{aligned}\chi(s) \equiv & 2(1 - \tau s) \left[A + \frac{B}{2}(s - \theta_{m+1}) \right] \prod_{j=1}^m (x_j + s) \\ & + (A + Bs)(z + s)(w + s) \sum_{i=1}^m \left[\prod_{j \neq i} (x_j + s) \cdot \frac{1 + \tau x_i}{x_i - \theta_{m+1}} \right]\end{aligned}\quad (\text{S54})$$

is a polynomial function of s . Note that the coefficient of s^{m+2} in (S54) is

$$\begin{aligned}-B\tau + B \sum_{i=1}^m \frac{1 + \tau x_i}{x_i - \theta_{m+1}} &= B(1 - \tau) \\ &= B \left(1 - \frac{\kappa + 1}{\kappa m} \right) \\ &= \frac{B}{\kappa m} [\kappa(m - 1) - 1] \\ &= \frac{B}{\kappa m} (mc^2 - 1) \neq 0,\end{aligned}$$

where the first equality follows from (S33), the second equality follows from the definition of τ , the last equality follows from the definition of κ , and the last inequality holds because $B = z + w - 2\theta_{m+1} > 0$ and $c \neq m^{-1/2}$. Thus, (S54) cannot be zero for all $s \in (0, |\theta_{m+1}|)$. This implies that $L(\tilde{h}(x, s)^2)$ must be positive. Consequently, $\frac{\partial^2 \mathbb{P}[|T_m| > c]}{\partial z^2}$ is also positive.

From the above, Lemma S1.11 holds. □

S1.4 Bounds on the derivative of the rejection probability

S1.4.1 More accurate bounds on the unique negative root θ_{m+1}

Lemma S1.12. *Consider the same setup and notations in Lemma S1.6. For any $x \geq 0$ and integer $1 \leq k \leq m$, define*

$$h(\theta; x, k) = -(k - 1)x - [m - x + (m + 1 - k)\tau x] \theta + \theta^2.$$

- (a) *For any $x \geq 0$ and $1 \leq k \leq m$, $h(\theta; x, k)$, viewed as a function of θ , has a unique positive root, which is denoted as $\theta(x, k)$.*
- (b) *$\theta(x, k) \geq m$ for any $x \geq 0$ and $1 \leq k \leq m$.*
- (c) *$\theta(0, k) = m$ for any $1 \leq k \leq m$, and $\theta(x, 1) = m + x/\kappa$ for any $x \geq 0$.*

(d) For any $1 \leq k \leq n$, $|\theta_{m+1}| \geq \theta(x_{(k)}, k) = \theta(\kappa\gamma_{(k)}^2, k)$, where $\gamma_{(1)} \leq \gamma_{(2)} \leq \dots \leq \gamma_{(m)}$ are the sorted values of $\{\gamma_j\}_{j=1}^m$, and $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(m)}$ are the sorted values of $\{x_j\}_{j=1}^m$.

Proof of Lemma S1.12(a)–(c). First, we consider the case when $x = 0$. In this case, $h(\theta; x, k)$ simplifies to $h(\theta; x, k) = -m\theta + \theta^2 = \theta(\theta - m)$. Obviously, $h(\theta; x, k)$ has a unique positive root at $\theta = m$.

Second, we consider the case when $k = 1$. In this case, $h(\theta; x, k)$ simplifies to $h(\theta; x, k) = -(m - x + m\tau x)\theta + \theta^2 = \theta\{\theta - m - (m\tau - 1)x\}$. Note that $m\tau - 1 = m \cdot (\frac{\kappa+1}{\kappa m}) - 1 = \frac{1}{\kappa}$ by definition. Consequently, $h(\theta; x, k) = \theta(\theta - m - \frac{x}{\kappa})$. Thus, $h(\theta; x, k)$ has a unique positive root at $\theta = m + \frac{x}{\kappa}$.

Third, we consider the case when $x > 0$ and $k > 1$. In this case, $h(0; x, k) = -(k - 1)x < 0$. Note that $h(\theta; x, k)$ converges to infinity as θ goes to positive or negative infinity. Thus, $h(\theta; x, k)$ must have one positive root and one negative root.

From the above, (a) and (c) hold. Below we prove (b). Note that, for any $x \geq 0$ and $1 \leq k \leq m$,

$$\begin{aligned} h(m; x, k) &= -(k - 1)x - [m - x + (m + 1 - k)\tau x]m + m^2 \\ &= -(k - 1)x + mx - (m + 1 - k)\tau mx \\ &= (m + 1 - k)x(1 - \tau m) \\ &= (m + 1 - k)x \left(1 - \frac{\kappa + 1}{\kappa}\right) \leq 0, \end{aligned}$$

where the last equality follows from the definition of τ . By the property of quadratic functions, we must have $\theta(x, k) \geq m$, i.e., (b) holds. \square

Proof of Lemma S1.12(d). From (S33), we know that

$$\begin{aligned} 1 &= \sum_{i=1}^m \frac{1 + \tau x_i}{x_i - \theta_{m+1}} = \sum_{i=1}^m \frac{1}{x_i - \theta_{m+1}} + \sum_{i=1}^m \frac{\tau x_i}{x_i - \theta_{m+1}} = \frac{1}{\theta_{m+1}} \sum_{i=1}^m \frac{\theta_{m+1}}{x_i - \theta_{m+1}} + \tau \sum_{i=1}^m \frac{x_i}{x_i - \theta_{m+1}} \\ &= \frac{1}{\theta_{m+1}} \sum_{i=1}^m \left(\frac{x_i}{x_i - \theta_{m+1}} - 1 \right) + \tau \sum_{i=1}^m \frac{x_i}{x_i - \theta_{m+1}} = \left(\frac{1}{\theta_{m+1}} + \tau \right) \sum_{i=1}^m \frac{x_i}{x_i - \theta_{m+1}} - \frac{m}{\theta_{m+1}}. \end{aligned} \tag{S55}$$

Consequently,

$$0 = -(m + \theta_{m+1}) + (1 + \tau\theta_{m+1}) \sum_{j=1}^m \frac{x_j}{x_j - \theta_{m+1}}.$$

From Lemma S1.4(c), $1 + \tau\theta_{m+1} < 0$. Because $x/(x - \theta_{m+1})$ is increasing in $x \geq 0$, we have, for any $1 \leq k \leq m$,

$$\begin{aligned} 0 &= -(m + \theta_{m+1}) + (1 + \tau\theta_{m+1}) \sum_{j=1}^m \frac{x_j}{x_j - \theta_{m+1}} \\ &= -(m + \theta_{m+1}) + (1 + \tau\theta_{m+1}) \sum_{j=1}^m \frac{x_{(j)}}{x_{(j)} - \theta_{m+1}} \\ &\leq -(m + \theta_{m+1}) + (1 + \tau\theta_{m+1})(m + 1 - k) \frac{x_{(k)}}{x_{(k)} - \theta_{m+1}}. \end{aligned}$$

Multiplying both left-hand and right-hand sides by $x_{(k)} - \theta_{m+1}$, we have

$$\begin{aligned} 0 &\leq -(m + \theta_{m+1})(x_{(k)} - \theta_{m+1}) + (1 + \tau\theta_{m+1})(m + 1 - k)x_{(k)} \\ &= -mx_{(k)} + m\theta_{m+1} - \theta_{m+1}x_{(k)} + \theta_{m+1}^2 + (m + 1 - k)x_{(k)} + (m + 1 - k)x_{(k)}\tau\theta_{m+1} \\ &= -(k - 1)x_{(k)} + \left[m - x_{(k)} + (m + 1 - k)\tau x_{(k)} \right] \theta_{m+1} + \theta_{m+1}^2 \\ &= -(k - 1)x_{(k)} - \left[m - x_{(k)} + (m + 1 - k)\tau x_{(k)} \right] |\theta_{m+1}| + |\theta_{m+1}|^2, \end{aligned}$$

i.e., $h(|\theta_{m+1}|; x, k) \geq 0$. From Lemma S1.12(a) and the properties of quadratic functions, we know that $|\theta_{m+1}| > 0$ must be no less than $\theta(x_{(k)}, k)$. \square

Lemma S1.13. Consider the same setup and notations in Lemmas S1.6 and S1.12. $|\theta_{m+1}| \leq m + x_{(m)}/\kappa = m + \gamma_{(m)}^2$, recalling that $x_{(m)} = \max_{1 \leq i \leq m} x_i$ and $\gamma_{(m)} = \max_{1 \leq i \leq m} \gamma_i$.

Proof of Lemma S1.13. From (S33) and by the same logic as the proof of Lemma S1.12,

$$0 = -(m + \theta_{m+1}) + (1 + \tau\theta_{m+1}) \sum_{j=1}^m \frac{x_j}{x_j - \theta_{m+1}}.$$

From Lemma S1.4, $\tau|\theta_{m+1}| > 1$. Because $\frac{x}{x + |\theta_{m+1}|}$ is increasing in x , we then have

$$|\theta_{m+1}| - m = (\tau|\theta_{m+1}| - 1) \sum_{j=1}^m \frac{x_j}{x_j + |\theta_{m+1}|} \leq (\tau|\theta_{m+1}| - 1) \frac{mx_{(m)}}{x_{(m)} + |\theta_{m+1}|},$$

which further implies that

$$\begin{aligned}
& (x_{(m)} + |\theta_{m+1}|)(|\theta_{m+1}| - m) \leq (\tau|\theta_{m+1}| - 1)mx_{(m)} \\
\implies & x_{(m)}|\theta_{m+1}| + |\theta_{m+1}|^2 - mx_{(m)} - m|\theta_{m+1}| \leq \tau|\theta_{m+1}|mx_{(m)} - mx_{(m)} \\
\implies & (x_{(m)} + |\theta_{m+1}| - m)|\theta_{m+1}| \leq \tau|\theta_{m+1}|mx_{(m)} \\
\implies & x_{(m)} + |\theta_{m+1}| - m \leq \tau mx_{(m)} \\
\implies & |\theta_{m+1}| \leq m + (\tau m - 1)x_{(m)} = m + \frac{x_{(m)}}{\kappa} = m + \gamma_{(m)}^2,
\end{aligned}$$

where the last two equalities follow from definition, recalling that $\tau = \frac{\kappa+1}{\kappa m}$. Therefore, Lemma S1.13 holds. \square

S1.4.2 Lemmas for bounding the derivative of the rejection probability

Lemma S1.14. *Consider the same setup and notations as in Lemma S1.6. Suppose that $z \equiv \max_{1 \leq i \leq m} x_i$ and $|\theta_{m+1}| \in [\underline{\theta}, \bar{\theta}]$ for some $\underline{\theta} \geq m$. For any given $v \in \mathbb{R}$, define*

$$G(s) \equiv \frac{s^v U(\mathbf{x}, s)}{(w + s)(z + s)}.$$

Then for any $s \in (0, |\theta_{m+1}|]$,

$$\frac{d \log G(s)}{ds} \geq \frac{1}{s} \left\{ v - \left[1 + \frac{\bar{\theta}}{2(z + \bar{\theta})} + \frac{\bar{\theta}}{w + \bar{\theta}} - \frac{\underline{\theta} - m}{2(\tau \underline{\theta} - 1)} \right] \right\}.$$

Proof of Lemma S1.14. Using the definition of $U(\mathbf{x}, s)$ in Lemma S1.6, we have

$$\begin{aligned}
& \log G(s) \\
&= \left(\frac{m}{2} - 1 + v \right) \log s - \frac{1}{2} \log P(\mathbf{x}, s) - \frac{1}{2} \log Q(\mathbf{x}, s) - \log(w + s) - \log(z + s).
\end{aligned} \tag{S56}$$

Since, by definition,

$$\frac{\partial P(\mathbf{x}, s)}{\partial s} = - \sum_{i=1}^m \frac{1 + \tau x_i}{(x_i + s)^2 (x_i - \theta_{m+1})},$$

we have

$$\frac{\partial \log P(\mathbf{x}, s)}{\partial s} = \frac{1}{P(\mathbf{x}, s)} \frac{\partial P(\mathbf{x}, s)}{\partial s} = - \frac{\sum_{i=1}^m \frac{1 + \tau x_i}{(x_i + s)^2 (x_i - \theta_{m+1})}}{\sum_{i=1}^m \frac{1 + \tau x_i}{(x_i + s)(x_i - \theta_{m+1})}}. \tag{S57}$$

Recall that $z = \max_{1 \leq i \leq m} x_i$. It follows that $\frac{1}{z+s} \leq \frac{1}{x_i+s}$ for any $1 \leq i \leq m$. Thus,

$$-\frac{\partial \log P(\mathbf{x}, s)}{\partial s} = \frac{\sum_{i=1}^m \frac{1+\tau x_i}{(x_i+s)^2(x_i-\theta_{m+1})}}{\sum_{i=1}^m \frac{1+\tau x_i}{(x_i+s)(x_i-\theta_{m+1})}} \geq \frac{\frac{1}{z+s} \sum_{i=1}^m \frac{1+\tau x_i}{(x_i+s)(x_i-\theta_{m+1})}}{\sum_{i=1}^m \frac{1+\tau x_i}{(x_i+s)(x_i-\theta_{m+1})}} = \frac{1}{z+s}. \quad (\text{S58})$$

Since, by definition, $\log Q(\mathbf{x}, s) = \sum_{i=1}^m \log(x_i + s)$, we have

$$\frac{\partial \log Q(\mathbf{x}, s)}{\partial s} = \sum_{i=1}^m \frac{1}{x_i + s}. \quad (\text{S59})$$

Combining the results from (S57)–(S59) into (S56), we have

$$\begin{aligned} \frac{d \log G(s)}{ds} &= \left(\frac{m}{2} - 1 + \nu\right) \frac{1}{s} - \frac{1}{2} \frac{\partial \log P(\mathbf{x}, s)}{\partial s} - \frac{1}{2} \frac{\partial \log Q(\mathbf{x}, s)}{\partial s} - \frac{1}{w+s} - \frac{1}{z+s} \\ &\geq \left(\frac{m}{2} - 1 + \nu\right) \frac{1}{s} + \frac{1}{2} \frac{1}{z+s} - \frac{1}{2} \sum_{i=1}^m \frac{1}{x_i + s} - \frac{1}{w+s} - \frac{1}{z+s} \\ &= \left(\frac{m}{2} - 1 + \nu\right) \frac{1}{s} - \frac{1}{2} \frac{1}{z+s} - \frac{1}{2} \sum_{i=1}^m \frac{1}{x_i + s} - \frac{1}{w+s} \\ &= \frac{1}{2} \sum_{i=1}^m \left(\frac{1}{s} - \frac{1}{x_i + s}\right) - \frac{1-\nu}{s} - \frac{1}{2(z+s)} - \frac{1}{w+s} \\ &= \frac{1}{2s} \sum_{i=1}^m \frac{x_i}{x_i + s} - \frac{1-\nu}{s} - \frac{1}{2(z+s)} - \frac{1}{w+s} \\ &= \frac{1}{2s} \left[\sum_{i=1}^m \frac{x_i}{x_i + s} - 2(1-\nu) - \frac{s}{z+s} - \frac{2s}{w+s} \right]. \end{aligned} \quad (\text{S60})$$

Below we bound the terms in (S60).

We first bound the first term in (S60). Because $s \leq |\theta_{m+1}| = -\theta_{m+1}$, we have

$$\sum_{i=1}^m \frac{x_i}{x_i + s} \geq \sum_{i=1}^m \frac{x_i}{x_i - \theta_{m+1}}.$$

Next, from (S33) and (S55), we know that

$$1 = \left(\frac{1}{\theta_{m+1}} + \tau\right) \sum_{i=1}^m \frac{x_i}{x_i - \theta_{m+1}} - \frac{m}{\theta_{m+1}}.$$

Consequently,

$$\begin{aligned}\sum_{i=1}^m \frac{x_i}{x_i + s} &\geq \sum_{i=1}^m \frac{x_i}{x_i - \theta_{m+1}} = \frac{1 + \frac{m}{\theta_{m+1}}}{\frac{1}{\theta_{m+1}} + \tau} = \frac{-\theta_{m+1} - m}{-\tau\theta_{m+1} - 1} = \frac{|\theta_{m+1}| - m}{\tau|\theta_{m+1}| - 1} \\ &= \tau^{-1} - \tau^{-1} \frac{m\tau - 1}{\tau|\theta_{m+1}| - 1}.\end{aligned}$$

Because, by definition, $m\tau = \frac{\kappa+1}{\kappa} > 1$, using the condition that $\tau|\theta_{m+1}| \geq \tau\theta \geq \tau m > 1$, we then have

$$\sum_{i=1}^m \frac{x_i}{x_i + s} \geq \tau^{-1} - \tau^{-1} \frac{m\tau - 1}{\tau|\theta_{m+1}| - 1} \geq \tau^{-1} - \tau^{-1} \frac{m\tau - 1}{\tau\theta - 1} = \frac{\theta - m}{\tau\theta - 1}.$$

We then bound the third term in (S60). Because $\frac{s}{z+s}$ is increasing in $s > 0$ and $s \leq |\theta_{m+1}| \leq \bar{\theta}$, we have

$$\frac{s}{z+s} \leq \frac{|\theta_{m+1}|}{z + |\theta_{m+1}|} \leq \frac{\bar{\theta}}{z + \bar{\theta}}.$$

We finally bound the last term in (S60). By the same logic as before, we have

$$\frac{s}{w+s} \leq \frac{|\theta_{m+1}|}{w + |\theta_{m+1}|} \leq \frac{\bar{\theta}}{w + \bar{\theta}}.$$

From the above, we have

$$\begin{aligned}\frac{d \log G(s)}{ds} &\geq \frac{1}{2s} \left[\sum_{i=1}^m \frac{x_i}{x_i + s} - 2(1 - \nu) - \frac{s}{z+s} - \frac{2s}{w+s} \right] \\ &\geq \frac{1}{2s} \left[\frac{\theta - m}{\tau\theta - 1} - 2(1 - \nu) - \frac{\bar{\theta}}{z + \bar{\theta}} - \frac{2\bar{\theta}}{w + \bar{\theta}} \right] \\ &= \frac{1}{s} \left[\frac{\theta - m}{2(\tau\theta - 1)} - 1 + \nu - \frac{\bar{\theta}}{2(z + \bar{\theta})} - \frac{\bar{\theta}}{w + \bar{\theta}} \right] \\ &= \frac{1}{s} \left\{ \nu - \left[1 + \frac{\bar{\theta}}{2(z + \bar{\theta})} + \frac{\bar{\theta}}{w + \bar{\theta}} - \frac{\theta - m}{2(\tau\theta - 1)} \right] \right\}.\end{aligned}$$

Therefore, Lemma S1.14 holds. □

Lemma S1.15. For any $\nu < 1$ and $\zeta > 0$,

$$\int_0^\zeta \frac{s^{-\nu}}{\sqrt{\zeta-s}} ds = \sqrt{\pi} \zeta^{\frac{1}{2}-\nu} \frac{\Gamma(1-\nu)}{\Gamma(\frac{3}{2}-\nu)}.$$

Proof of Lemma S1.15. To begin with, using the properties of beta and gamma functions,

$$\int_0^1 \frac{x^{-\nu}}{\sqrt{1-x}} dx = \frac{\Gamma(1-\nu)\Gamma(\frac{1}{2})}{\Gamma(\frac{3}{2}-\nu)} = \frac{\Gamma(1-\nu)\sqrt{\pi}}{\Gamma(\frac{3}{2}-\nu)}. \quad (\text{S61})$$

The integral (S61) is convergent if $1-\nu > 0$, or equivalently, $\nu < 1$. By change of variables, let $s = \zeta x$. Hence, $ds = \zeta dx$. Therefore,

$$\int_0^\zeta \frac{s^{-\nu}}{\sqrt{\zeta-s}} ds = \int_0^1 \frac{\zeta^{-\nu} x^{-\nu}}{\sqrt{\zeta-\zeta x}} \zeta dx = \zeta^{\frac{1}{2}-\nu} \int_0^1 \frac{x^{-\nu}}{\sqrt{1-x}} dx = \sqrt{\pi} \zeta^{\frac{1}{2}-\nu} \frac{\Gamma(1-\nu)}{\Gamma(\frac{3}{2}-\nu)},$$

where the last equality uses (S61). Therefore, Lemma S1.15 holds. \square

Lemma S1.16. Consider the same setup and notations in Lemma S1.6, and adopt the notations from Lemmas S1.7 and S1.8. Suppose that $z \equiv \max_{1 \leq i \leq m} x_i > 0$, $w \equiv \min_{1 \leq i \leq m} x_i$, $|\theta_{m+1}| \leq \bar{\theta}$ for some positive $\bar{\theta}$, and $\tau \leq 1/2$. Define

$$F(s) \equiv \frac{s^{-\nu}(w+s)(z+s)\tilde{h}(x,s)}{\sqrt{|\theta_{m+1}|-s}},$$

and

$$C_1 \equiv (1-2\tau) \frac{wz - \bar{\theta}^2}{(w+z+2\bar{\theta})\bar{\theta}} + \frac{2 \frac{wz - \bar{\theta}^2}{(w+z+2\bar{\theta})\bar{\theta}} + 1}{z} - \tau.$$

(a) For any $s \in (0, |\theta_{m+1}|)$,

$$F(s) \geq \frac{Bs^{-\nu}}{\sqrt{|\theta_{m+1}|-s}} [|\theta_{m+1}|C_1 + (1-\tau)s].$$

(b) If

$$\nu < \min \left\{ \frac{3}{2} - \frac{1}{2(1 + \frac{C_1}{1-\tau})}, 1 \right\}$$

then

$$\int_0^{|\theta_{m+1}|} \frac{s^{-\nu}}{\sqrt{|\theta_{m+1}| - s}} [|\theta_{m+1}|C_1 + (1 - \tau)s] ds > 0.$$

Proof of Lemma S1.16(a). From the definition of $\tilde{h}(\mathbf{x}, s)$ in Lemma S1.8, we have

$$(w + s)(z + s)\tilde{h}(\mathbf{x}, s) = A + Bs + \frac{2(A + \frac{s - \theta_{m+1}}{2}B)}{(w + s)(z + s)P(\mathbf{x}, s)} - \frac{2\tau s(A + \frac{s - \theta_{m+1}}{2}B)}{(w + s)(z + s)P(\mathbf{x}, s)}. \quad (\text{S62})$$

In the following, we bound (S62) from below.

First, because $z = \max_{1 \leq i \leq m} x_i$, we have

$$(w + s)(z + s)P(\mathbf{x}, s) = (w + s)(z + s) \sum_{i=1}^m \frac{1 + \tau x_i}{(x_i + s)(x_i - \theta_{m+1})} \quad (\text{S63})$$

$$\begin{aligned} &= (w + s) \sum_{i=1}^m \left(\frac{z + s}{x_i + s} \cdot \frac{1 + \tau x_i}{x_i - \theta_{m+1}} \right) \\ &\geq (w + s) \sum_{i=1}^m \frac{1 + \tau x_i}{x_i - \theta_{m+1}} \\ &= w + s \geq s, \end{aligned} \quad (\text{S64})$$

where the last equality follows from (S33) and the last inequality follows from the fact that $w \geq 0$.

Second, because $w = \min_{1 \leq i \leq m} x_i$, we have

$$\begin{aligned} (w + s)(z + s)P(\mathbf{x}, s) &= (w + s)(z + s) \sum_{i=1}^m \frac{1 + \tau x_i}{(x_i + s)(x_i - \theta_{m+1})} \\ &= (z + s) \sum_{i=1}^m \left(\frac{w + s}{x_i + s} \cdot \frac{1 + \tau x_i}{x_i - \theta_{m+1}} \right) \\ &\leq (z + s) \sum_{i=1}^m \frac{1 + \tau x_i}{x_i - \theta_{m+1}} = z + s, \end{aligned} \quad (\text{S65})$$

where the last equality follows from (S33).

Third, from the definition of A and B in Lemma S1.7, we know that

$$\begin{aligned} A + \frac{s - \theta_{m+1}}{2}B &= wz - \theta_{m+1}^2 + \frac{s - \theta_{m+1}}{2}(z + w - 2\theta_{m+1}) \\ &= wz - |\theta_{m+1}|^2 + \frac{s + |\theta_{m+1}|}{2}(z + w + 2|\theta_{m+1}|) \end{aligned}$$

$$\begin{aligned}
&= s \left(\frac{z+w}{2} + |\theta_{m+1}| \right) + wz - |\theta_{m+1}|^2 + |\theta_{m+1}|^2 + |\theta_{m+1}| \frac{z+w}{2} \\
&= s \left(\frac{z+w}{2} + |\theta_{m+1}| \right) + z \left(w + \frac{|\theta_{m+1}|}{2} \right) + \frac{w|\theta_{m+1}|}{2} \\
&\geq 0.
\end{aligned} \tag{S66}$$

In addition, because $\frac{A + \frac{s-\theta_{m+1}}{2}B}{z+s} = \frac{(A + \frac{|\theta_{m+1}|}{2}B) + \frac{B}{2}s}{z+s}$ is monotone in s , we have, for any $s \in [0, |\theta_{m+1}|]$,

$$\begin{aligned}
\frac{A + \frac{s-\theta_{m+1}}{2}B}{z+s} &\geq \min \left\{ \frac{A + \frac{|\theta_{m+1}|}{2}B}{z}, \frac{A + |\theta_{m+1}|B}{z + |\theta_{m+1}|} \right\} \\
&\geq \min \left\{ \frac{A + \frac{|\theta_{m+1}|}{2}B}{z}, \frac{A + |\theta_{m+1}|B}{z + \bar{\theta}} \right\},
\end{aligned} \tag{S67}$$

where the last inequality holds due to (S66) and that $|\theta_{m+1}| \leq \bar{\theta}$.

Fourth, from the previous three parts, we have

$$\begin{aligned}
& (w+s)(z+s)\tilde{h}(\mathbf{x}, s) \\
&= A + Bs + \frac{2(A + \frac{s-\theta_{m+1}}{2}B)}{(w+s)(z+s)P(\mathbf{x}, s)} - \frac{2\tau s(A + \frac{s-\theta_{m+1}}{2}B)}{(w+s)(z+s)P(\mathbf{x}, s)} \\
&\geq A + Bs + \frac{2(A + \frac{s-\theta_{m+1}}{2}B)}{z+s} - 2\tau \left(A + \frac{s-\theta_{m+1}}{2}B \right) \quad \text{using (S64), (S65), and (S66)} \\
&\geq A + Bs + 2 \cdot \min \left\{ \frac{A + \frac{|\theta_{m+1}|}{2}B}{z}, \frac{A + |\theta_{m+1}|B}{z + \bar{\theta}} \right\} - 2\tau \left(A + \frac{s-\theta_{m+1}}{2}B \right) \quad \text{using (S67)} \\
&= (1-2\tau)A + 2 \cdot \min \left\{ \frac{A + \frac{|\theta_{m+1}|}{2}B}{z}, \frac{A + |\theta_{m+1}|B}{z + \bar{\theta}} \right\} - \tau|\theta_{m+1}|B + (1-\tau)Bs \\
&= |\theta_{m+1}|B \cdot \left[(1-2\tau) \frac{A}{B|\theta_{m+1}|} + 2 \cdot \min \left\{ \frac{\frac{A}{B|\theta_{m+1}|} + \frac{1}{2}}{z}, \frac{\frac{A}{B|\theta_{m+1}|} + 1}{z + \bar{\theta}} \right\} - \tau \right] + (1-\tau)Bs.
\end{aligned} \tag{S68}$$

Note that

$$\begin{aligned}
\frac{A}{B|\theta_{m+1}|} &= \frac{wz - |\theta_{m+1}|^2}{(w+z+2|\theta_{m+1}|)|\theta_{m+1}|} = \frac{wz}{(w+z+2|\theta_{m+1}|)|\theta_{m+1}|} - \frac{|\theta_{m+1}|}{w+z+2|\theta_{m+1}|} \\
&\geq \frac{wz}{(w+z+2\bar{\theta})\bar{\theta}} - \frac{\bar{\theta}}{w+z+2\bar{\theta}} = \frac{wz - \bar{\theta}^2}{(w+z+2\bar{\theta})\bar{\theta}}.
\end{aligned}$$

Because $\tau \leq 1/2$, we then have

$$\begin{aligned}
& (1 - 2\tau) \frac{A}{B|\theta_{m+1}|} + 2 \cdot \min \left\{ \frac{\frac{A}{B|\theta_{m+1}|} + \frac{1}{2}}{z}, \frac{\frac{A}{B|\theta_{m+1}|} + 1}{z + \bar{\theta}} \right\} - \tau \\
& \geq (1 - 2\tau) \frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + 2 \cdot \min \left\{ \frac{\frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + \frac{1}{2}}{z}, \frac{\frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + 1}{z + \bar{\theta}} \right\} - \tau.
\end{aligned}$$

Note that

$$\begin{aligned}
& \frac{\frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + \frac{1}{2}}{z} \leq \frac{\frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + 1}{z + \bar{\theta}} \\
& \iff \frac{(wz - \bar{\theta}^2)z}{(w + z + 2\bar{\theta})\bar{\theta}} + \frac{z}{2} + \frac{(wz - \bar{\theta}^2)\bar{\theta}}{(w + z + 2\bar{\theta})\bar{\theta}} + \frac{\bar{\theta}}{2} \leq \frac{(wz - \bar{\theta}^2)z}{(w + z + 2\bar{\theta})\bar{\theta}} + z \\
& \iff \frac{z}{2} \geq \frac{wz - \bar{\theta}^2}{w + z + 2\bar{\theta}} + \frac{\bar{\theta}}{2} = \frac{wz - \bar{\theta}^2 + (w + z + 2\bar{\theta})\frac{\bar{\theta}}{2}}{w + z + 2\bar{\theta}} = \frac{wz + (w + z)\frac{\bar{\theta}}{2}}{w + z + 2\bar{\theta}} \\
& \iff (w + z)z + 2z\bar{\theta} \geq 2wz + (w + z)\bar{\theta} \\
& \iff z(z - w) + \bar{\theta}(z - w) \geq 0,
\end{aligned}$$

which must hold because $z \geq w \geq 0$ and $\bar{\theta} \geq 0$. Thus, we have

$$\begin{aligned}
& (1 - 2\tau) \frac{A}{B|\theta_{m+1}|} + 2 \cdot \min \left\{ \frac{\frac{A}{B|\theta_{m+1}|} + \frac{1}{2}}{z}, \frac{\frac{A}{B|\theta_{m+1}|} + 1}{z + \bar{\theta}} \right\} - \tau \\
& \geq (1 - 2\tau) \frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + \frac{2 \frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + 1}{z} - \tau \\
& = C_1,
\end{aligned}$$

where the last equality holds by definition. From (S68), we then have

$$(w + s)(z + s)\tilde{h}(\mathbf{x}, s) \geq |\theta_{m+1}|BC_1 + (1 - \tau)Bs$$

This immediately implies that

$$\begin{aligned}
F(s) & \equiv \frac{s^{-\nu}(w + s)(z + s)\tilde{h}(\mathbf{x}, s)}{\sqrt{|\theta_{m+1}|} - s} \geq \frac{s^{-\nu}}{\sqrt{|\theta_{m+1}|} - s} [|\theta_{m+1}|BC_1 + (1 - \tau)Bs] \\
& = \frac{Bs^{-\nu}}{\sqrt{|\theta_{m+1}|} - s} [|\theta_{m+1}|C_1 + (1 - \tau)s].
\end{aligned}$$

From the above, Lemma S1.16(a) holds. □

Proof of Lemma S1.16(b). From Lemma S1.15, for $\nu < 1$, we have

$$\begin{aligned}
& \int_0^{|\theta_{m+1}|} \frac{s^{-\nu}}{\sqrt{|\theta_{m+1}| - s}} [|\theta_{m+1}|C_1 + (1 - \tau)s] \, ds \\
&= |\theta_{m+1}|C_1 \cdot \int_0^{|\theta_{m+1}|} \frac{s^{-\nu}}{\sqrt{|\theta_{m+1}| - s}} \, ds + (1 - \tau) \int_0^{|\theta_{m+1}|} \frac{s^{-(\nu-1)}}{\sqrt{|\theta_{m+1}| - s}} \, ds \\
&= C_1 \sqrt{\pi} |\theta_{m+1}|^{\frac{3}{2}-\nu} \frac{\Gamma(1-\nu)}{\Gamma(\frac{3}{2}-\nu)} + (1 - \tau) \sqrt{\pi} |\theta_{m+1}|^{\frac{3}{2}-\nu} \frac{\Gamma(2-\nu)}{\Gamma(\frac{5}{2}-\nu)} \\
&= \sqrt{\pi} |\theta_{m+1}|^{\frac{3}{2}-\nu} \frac{\Gamma(1-\nu)}{\Gamma(\frac{3}{2}-\nu)} \left[C_1 + (1 - \tau) \frac{1-\nu}{\frac{3}{2}-\nu} \right],
\end{aligned}$$

where the last equality uses the property of gamma functions. Note that, by definition,

$$\frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + \frac{1}{2} = \frac{wz + \frac{1}{2}(w + z)\bar{\theta}}{(w + z + 2\bar{\theta})\bar{\theta}} > 0, \tag{S69}$$

which immediately implies that

$$\begin{aligned}
C_1 + (1 - \tau) &= (1 - 2\tau) \frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + \frac{2 \frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + 1}{z} - \tau + (1 - \tau) \\
&= (1 - 2\tau) \left[\frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + 1 \right] + \frac{2 \frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + 1}{z} > 0.
\end{aligned}$$

Consequently, when $\nu < 1$ and $\tau < \frac{1}{2}$, we have

$$\begin{aligned}
& C_1 + (1 - \tau) \frac{1-\nu}{\frac{3}{2}-\nu} > 0 \\
& \iff C_1 + (1 - \tau) \left(1 - \frac{\frac{1}{2}}{\frac{3}{2}-\nu} \right) > 0 \\
& \iff \frac{C_1}{1 - \tau} + 1 > \frac{\frac{1}{2}}{\frac{3}{2}-\nu} \\
& \iff \frac{3}{2} - \nu > \frac{1}{2(1 + \frac{C_1}{1-\tau})} \\
& \iff \nu < \frac{3}{2} - \frac{1}{2(1 + \frac{C_1}{1-\tau})},
\end{aligned}$$

where the second last equivalence holds because $\frac{C_1}{1-\tau} + 1 = \frac{C_1+1-\tau}{1-\tau} > 0$. Therefore, Lemma S1.16(b) holds. \square

Lemma S1.17. *Consider the same setup and notations in Lemma S1.6, and adopt the notations from Lemmas S1.7 and S1.8. Suppose that $z \equiv \max_{1 \leq i \leq m} x_i > 0$, $|\theta_{m+1}| \leq \bar{\theta}$ for some positive $\bar{\theta}$, and $\tau \leq 1/2$. Define*

$$F(s) \equiv \frac{s^{-\nu}(w+s)(z+s)\tilde{h}(\mathbf{x}, s)}{\sqrt{|\theta_{m+1}| - s}},$$

and

$$C_2 \equiv (1 - 2\tau) \frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} - \tau.$$

(a) For any $s \in (0, |\theta_{m+1}|)$,

$$F(s) \geq \frac{Bs^{-\nu}}{\sqrt{|\theta_{m+1}| - s}} [|\theta_{m+1}|C_2 + (1 - \tau)s].$$

(b) If

$$\nu < \min \left\{ \frac{3}{2} - \frac{1}{2(1 + \frac{C_2}{1-\tau})}, 1 \right\}$$

then

$$\int_0^{|\theta_{m+1}|} \frac{s^{-\nu}}{\sqrt{|\theta_{m+1}| - s}} [|\theta_{m+1}|C_2 + (1 - \tau)s] ds > 0.$$

Proof of Lemma S1.17. The proof of Lemma S1.17 follows almost the same steps as the proof of Lemma S1.16, except for a few differences discussed below. For (a), we no longer have (S65) and instead bound the second term in (S62) by zero. That is, $\frac{2(A + \frac{s-\theta_{m+1}}{2}B)}{(w+s)(z+s)P(\mathbf{x}, s)} \geq 0$, which follows immediately from (S66). For (b), we also have

$$\begin{aligned} C_2 + (1 - \tau) &= (1 - 2\tau) \frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} - \tau + (1 - \tau) \\ &= (1 - 2\tau) \left[\frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + 1 \right] > 0, \end{aligned}$$

which follows immediately from (S69). □

S1.4.3 Bounds on the derivative of the rejection probability

Lemma S1.18 (Modified from Lemma 1 of Bakirov (1989)). *Let F and G be two functions of x that satisfy the following properties, where θ_{m+1} can be any given negative number in \mathbb{R} :*

(i) $G(x)$ is not identically equal to zero, continuous and nonnegative on $(0, |\theta_{m+1}|]$ and $\frac{d}{dx}G(x)$ is continuous and nonnegative on $(0, |\theta_{m+1}|)$.

(ii) $F(x)$ is continuous on $(0, |\theta_{m+1}|)$, and there exists some $x_0 < |\theta_{m+1}|$ such that $F(x)(x - x_0) \geq 0$ and $F(x) > 0$ for $x_0 < x < |\theta_{m+1}|$.

If $\int_0^{|\theta_{m+1}|} F(x) dx$ converges and is positive, then $\int_0^{|\theta_{m+1}|} F(x)G(x) dx$ converges, and

$$\int_0^{|\theta_{m+1}|} F(x)G(x) dx > 0.$$

Proof of Lemma S1.18. The result mainly follows from the proof of Lemma 1 in Bakirov (1989), with 1 in the integral limit replaced by $|\theta_{m+1}|$ and slightly revised conditions. For completeness purposes, we prove the lemma with the revised conditions below, although most of the proof are taken from Bakirov (1989).

First, from condition (i), $G(\cdot)$ is nonnegative and monotone nondecreasing on $(0, |\theta_{m+1}|]$. Thus, $G(x)$ must have a nonnegative limit as $x \rightarrow 0+$. Define the value of $G(\cdot)$ evaluated at 0 as this limit, i.e., $G(0) \equiv \lim_{x \rightarrow 0+} G(x) \geq 0$. Consequently, $G(\cdot)$ becomes a continuous and nonnegative function on $[0, |\theta_{m+1}|]$. In addition, there must exist a finite M such that $0 \leq G(x) \leq M$ for all $x \in [0, |\theta_{m+1}|]$.

Second, we prove that $\bar{F}(x) \equiv \int_x^{|\theta_{m+1}|} F(s) ds > 0$ for any $x \in [0, |\theta_{m+1}|)$. If condition (ii) holds for some $x_0 \leq 0$, then this holds obviously. Below we consider only the case where condition (ii) holds for some $x_0 \in (0, |\theta_{m+1}|)$. For any $x \geq x_0$, we obviously have $\int_x^{|\theta_{m+1}|} F(s) ds > 0$, since $F(x) > 0$ for $x > x_0$. For any $x < x_0$, we have

$$\int_x^{|\theta_{m+1}|} F(s) ds = \int_0^{|\theta_{m+1}|} F(s) ds - \int_0^x F(s) ds \geq \int_0^{|\theta_{m+1}|} F(s) ds > 0,$$

where the second last inequality holds because $F(s) \leq 0$ for $0 < s \leq x < x_0$.

Third, we prove that $\int_0^{|\theta_{m+1}|} F(x)G(x) dx$ converges. From condition (ii), there exists $0 < x_1 < |\theta_{m+1}|$ such that $F(x) \geq 0$ for all $x \in [x_1, |\theta_{m+1}|)$. From the first part, for any $x \in [x_1, |\theta_{m+1}|)$, we then have $0 \leq F(x)G(x) \leq MF(x)$. This implies that $\int_{x_1}^{|\theta_{m+1}|} F(x)G(x) dx$ is bounded from the above by $M \int_{x_1}^{|\theta_{m+1}|} F(x) dx$ and thus converges. We consider then

two cases depending on the value of x_0 in condition (ii).

- Consider first the case where condition (ii) holds for some $x_0 \leq 0$. Let x_2 be any number in $(0, |\theta_{m+1}|)$. We then have $F(x) \geq 0$ for all $x \in (0, x_2]$. From the first part, for any $x \in (0, x_2]$, we then have $0 \leq F(x)G(x) \leq MF(x)$. This implies that $\int_0^{x_2} F(x)G(x) dx$ is bounded from the above by $M \int_0^{x_2} F(x) dx$ and thus converges.
- Consider then the case where condition (ii) holds for some $x_0 > 0$. Let $x_3 < x_0$ be a number in $(0, |\theta_{m+1}|)$. We then have $F(x) \leq 0$ for all $x \in (0, x_3]$. From the first part, for any $x \in (0, x_3]$, we then have $MF(x) \leq F(x)G(x) \leq 0$. This implies that $\int_0^{x_3} F(x)G(x) dx$ is bounded from the below by $M \int_0^{x_3} F(x) dx$ and thus converges.

The above discussion then implies that $\int_0^{|\theta_{m+1}|} F(x)G(x) dx$ converges.

Fourth, recalling the definition of $\bar{F}(x)$ in the second part, we have $\frac{d}{dx}\bar{F}(x) = -F(x)$. For any positive ϵ_1 and ϵ_2 such that $\epsilon_1 + \epsilon_2 < |\theta_{m+1}|$, using integration by parts, we have

$$\begin{aligned}
& \int_{\epsilon_1}^{|\theta_{m+1}|-\epsilon_2} F(x)G(x) dx \\
&= - \int_{\epsilon_1}^{|\theta_{m+1}|-\epsilon_2} G(x) d\bar{F}(x) \\
&= - G(x)\bar{F}(x) \Big|_{\epsilon_1}^{|\theta_{m+1}|-\epsilon_2} + \int_{\epsilon_1}^{|\theta_{m+1}|-\epsilon_2} \bar{F}(x) dG(x) \\
&= -G(|\theta_{m+1}| - \epsilon_2)\bar{F}(|\theta_{m+1}| - \epsilon_2) + G(\epsilon_1)\bar{F}(\epsilon_1) + \int_{\epsilon_1}^{|\theta_{m+1}|-\epsilon_2} \frac{d}{dx}G(x)\bar{F}(x) dx. \quad (S70)
\end{aligned}$$

From the first part and the condition that $\int_0^{|\theta_{m+1}|} F(x) dx$ converges, we can know that

$$G(|\theta_{m+1}| - \epsilon_2)\bar{F}(|\theta_{m+1}| - \epsilon_2) = G(|\theta_{m+1}| - \epsilon_2) \int_{|\theta_{m+1}|-\epsilon_2}^{|\theta_{m+1}|} F(s) ds \longrightarrow G(|\theta_{m+1}|) \cdot 0 = 0$$

as $\epsilon_2 \longrightarrow 0+$, and

$$G(\epsilon_1)\bar{F}(\epsilon_1) \longrightarrow G(0) \int_0^{|\theta_{m+1}|} F(x) dx = G(0)\bar{F}(0)$$

as $\epsilon_1 \longrightarrow 0+$. From the third part, $\int_0^{|\theta_{m+1}|} F(x)G(x) dx$ converges. These imply that $\int_0^{|\theta_{m+1}|} \frac{d}{dx}G(x)\bar{F}(x) dx$ converges. By letting ϵ_1 and ϵ_2 in (S70) converge to zero from the right, we then have

$$\int_0^{|\theta_{m+1}|} F(x)G(x) dx = G(0)\bar{F}(0) + \int_0^{|\theta_{m+1}|} \frac{d}{dx}G(x)\bar{F}(x) dx. \quad (S71)$$

From the second part, we know that $\bar{F}(x) > 0$ for any $x \in [0, |\theta_{m+1}|)$. From condition (i) and the first part, we know that $G(0) \geq 0$ and $\frac{d}{dx}G(x)$ is continuous and nonnegative for any $x \in (0, |\theta_{m+1}|)$. Thus, the right-hand side of (S71) nonnegative, and it becomes zero if and only if $G(0) = 0$ and $\frac{d}{dx}G(x) = 0$ for any $x \in (0, |\theta_{m+1}|)$, under which $G(\cdot)$ becomes a zero function on $[0, |\theta_{m+1}|]$. Because $G(x)$ is not identically equal to zero as in condition (i), the right-hand side of (S71) must be positive. Therefore, $\int_0^{|\theta_{m+1}|} F(x)G(x) dx > 0$.

From the above, Lemma S1.18 holds. \square

Lemma S1.19. Consider the same setup and notations in Lemmas S1.6–S1.9. Let $z = x_{(m)}$, w be any one of (x_1, x_2, \dots, x_m) , $\bar{\theta} = m + z/\kappa$, $\underline{\theta} = \max_{1 \leq k \leq m} \theta(x_{(k)}, k)$, and

$$C \equiv \begin{cases} (1 - 2\tau) \frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + \frac{2 \frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} + 1}{z} - \tau, & \text{if } w = x_{(1)}, \\ (1 - 2\tau) \frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} - \tau, & \text{otherwise.} \end{cases}$$

Fix the value of c , the values of $\{x_i\}_{i=1}^m$ excluding (z, w) , and the value of θ_{m+1} , and view $\mathbb{P}[|T_m| > c]$ as a function of z . If $\tau \leq 1/2$, $z > w > 0$, and

$$\frac{\bar{\theta}}{z + \bar{\theta}} + \frac{2\bar{\theta}}{w + \bar{\theta}} - \frac{\underline{\theta} - m}{\tau \underline{\theta} - 1} + \frac{1 - \tau}{1 - \tau + \min\{C, 0\}} < 1$$

then $\frac{\partial \mathbb{P}[|T_m| > c]}{\partial z} < 0$.

Proof of Lemma S1.19. Let

$$\nu = 1 + \frac{\bar{\theta}}{2(z + \bar{\theta})} + \frac{\bar{\theta}}{w + \bar{\theta}} - \frac{\underline{\theta} - m}{2(\tau \underline{\theta} - 1)}, \quad (\text{S72})$$

and define $G(s)$ and $F(s)$ in the same way as in Lemmas S1.14–S1.17:

$$G(s) = \frac{s^\nu U(\mathbf{x}, s)}{(w + s)(z + s)},$$

$$F(s) = \frac{s^{-\nu}(w + s)(z + s)\tilde{h}(\mathbf{x}, s)}{\sqrt{|\theta_{m+1}| - s}}$$

for $s \in (0, |\theta_{m+1}|)$. From Lemma S1.11, we have

$$\frac{\partial \mathbb{P}[|T_m| > c]}{\partial z} = -\frac{\Delta}{2} L(\tilde{h}(\mathbf{x}, s)) = -\frac{\Delta}{2} \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{\tilde{h}(\mathbf{x}, s) U(\mathbf{x}, s)}{\sqrt{|\theta_{m+1}| - s}} ds$$

$$\begin{aligned}
&= -\frac{z-w}{2(z-\theta_{m+1})^2} \frac{1}{\pi} \int_0^{|\theta_{m+1}|} \frac{s^\nu U(\mathbf{x}, s)}{(w+s)(z+s)} \frac{s^{-\nu}(w+s)(z+s)\tilde{h}(\mathbf{x}, s)}{\sqrt{|\theta_{m+1}|-s}} ds \\
&= -\frac{z-w}{2(z-\theta_{m+1})^2} \frac{1}{\pi} \int_0^{|\theta_{m+1}|} G(s)F(s) ds.
\end{aligned}$$

Note that $z > w$. To prove Lemma S1.19, it suffices to prove that $\int_0^{|\theta_{m+1}|} G(s)F(s) ds > 0$ under the conditions in Lemma S1.19.

First, from Lemma S1.12, $\underline{\theta} = \max_{1 \leq k \leq m} \theta(x_{(k)}, k) \geq m$. From Lemmas S1.12 and S1.13, we must have $m \leq \underline{\theta} \leq |\theta_{m+1}| \leq \bar{\theta}$.

Second, by definition, we can know that $G(s)$ is positive and continuous on $(0, |\theta_{m+1}|]$. Moreover, by the definition in (S72) and using Lemma S1.14, we can know that $\frac{d \log G(s)}{ds} \geq 0$ for $s \in (0, |\theta_{m+1}|]$. These then imply that $\frac{dG(s)}{ds} \geq 0$ for $s \in (0, |\theta_{m+1}|]$.

Third, by the definition of C and using Lemmas S1.16 and S1.17 for $s \in (0, |\theta_{m+1}|)$, we have $F(s) \geq B\tilde{F}(s)$, where

$$\tilde{F}(s) \equiv \frac{s^{-\nu}}{\sqrt{|\theta_{m+1}|-s}} [|\theta_{m+1}|C + (1-\tau)s].$$

By the definition in (S72) and the conditions in Lemma S1.19, we have

$$\begin{aligned}
&2(\nu-1) + \frac{1-\tau}{1-\tau + \min\{C, 0\}} < 1 \\
\Rightarrow \quad \nu &< \frac{3}{2} - \frac{1}{2(1 + \frac{\min\{C, 0\}}{1-\tau})} = \min \left\{ \frac{3}{2} - \frac{1}{2(1 + \frac{C}{1-\tau})}, 1 \right\}.
\end{aligned}$$

From Lemmas S1.16 and S1.17, we then have $\int_0^{|\theta_{m+1}|} \tilde{F}(s) ds > 0$. Moreover, because $1-\tau > 0$, there must exists $s_0 < |\theta_{m+1}|$ such that $\tilde{F}(s) \cdot (s-s_0) \geq 0$ for $s \in (0, |\theta_{m+1}|)$ and $F(s) > 0$ for $s_0 < s < |\theta_{m+1}|$.

From the above, Lemma S1.18 and the fact that $B > 0$ from its definition, we then have

$$\int_0^{|\theta_{m+1}|} G(s)F(s) ds \geq B \cdot \int_0^{|\theta_{m+1}|} G(s)\tilde{F}(s) ds > 0.$$

Therefore, Lemma S1.19 holds. □

S2 Proof of theorems and lemmas

S2.1 Proof of Theorem 4.1

To prove Theorem 4.1, we need the following lemma.

Lemma S2.1. *Under Assumption 4.1, for any $\delta \in \mathbb{R}$ and $c > 0$, if $c \neq m^{-1/2}$ and $\{\sigma_j^2\}_{j=1}^{m+1}$ are not all zero, then $\mathbb{P}[(\psi_{m+1} - \bar{\psi}_m + \delta)^2 = c^2 S_m^2] = 0$.*

Proof of Lemma S2.1. If $\sigma_{m+1} > 0$, then, by the law of iterated expectation, we must have

$$\begin{aligned} \mathbb{P}[(\psi_{m+1} - \bar{\psi}_m + \delta)^2 = c^2 S_m^2] &= \mathbb{E} \left\{ \mathbb{P} \left[(\psi_{m+1} - \bar{\psi}_m + \delta)^2 = c^2 S_m^2 \mid \psi_1, \dots, \psi_m \right] \right\} \\ &= 0, \end{aligned} \tag{S73}$$

where the last equality holds because ψ_{m+1} is a continuous random variable and has zero probability mass at any real value. Below it suffices to consider the case where $\sigma_{m+1} = 0$.

From the proof of Lemma S1.1, $mc^2 S_m^2 - m(\psi_{m+1} - \bar{\psi}_m)^2 = \tilde{\psi}^\top \mathbf{V} \tilde{\psi}$, and $\psi_{m+1} - \bar{\psi}_m = \mathbf{a}^\top \tilde{\psi}$, where \mathbf{V} and $\tilde{\psi}$ are defined the same as in Lemma S1.1, and $\mathbf{a}^\top = (1, -m^{-1} \mathbf{1}_m^\top)$. Consequently,

$$\begin{aligned} \mathbb{P}[(\psi_{m+1} - \bar{\psi}_m + \delta)^2 = c^2 S_m^2] &= \mathbb{P}[m(\psi_{m+1} - \bar{\psi}_m + \delta)^2 = mc^2 S_m^2] \\ &= \mathbb{P}[mc^2 S_m^2 - m(\psi_{m+1} - \bar{\psi}_m)^2 - 2m\delta(\psi_{m+1} - \bar{\psi}_m) - m\delta^2 = 0] \\ &= \mathbb{P}[\tilde{\psi}^\top \mathbf{V} \tilde{\psi} - 2m\delta \mathbf{a}^\top \tilde{\psi} - m\delta^2 = 0] \\ &= \mathbb{P}[\boldsymbol{\xi}^\top \mathbf{D} \mathbf{V} \mathbf{D} \boldsymbol{\xi} - 2m\delta \mathbf{a}^\top \mathbf{D} \boldsymbol{\xi} - m\delta^2 = 0], \end{aligned}$$

where \mathbf{D} is defined as in Lemma S1.2, $\boldsymbol{\xi} \equiv (\xi_1, \xi_2, \dots, \xi_{m+1})^\top$, and $\{\xi_i\}_{i=1}^{m+1}$ are i.i.d. standard normal random variables. Let $\mathbf{D} \mathbf{V} \mathbf{D} = \boldsymbol{\Gamma} \boldsymbol{\Lambda} \boldsymbol{\Gamma}^\top$ be the eigendecomposition of $\mathbf{D} \mathbf{V} \mathbf{D}$, where $\boldsymbol{\Gamma}$ is an orthogonal matrix, $\boldsymbol{\Lambda}$ is a diagonal matrix with diagonal elements $\{\lambda_i\}_{i=1}^{m+1}$, and $\{\lambda_i\}_{i=1}^{m+1}$ are the eigenvalues of $\mathbf{D} \mathbf{V} \mathbf{D}$, or equivalently the root of the characteristic polynomial $f(\lambda)$ in Lemma S1.2. Let $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_{m+1})^\top = \boldsymbol{\Gamma}^\top \boldsymbol{\xi}$. We can verify that $\{\zeta_i\}_{i=1}^{m+1}$ are i.i.d. standard normal random variables, and $\boldsymbol{\xi} = \boldsymbol{\Gamma} \boldsymbol{\zeta}$. We then have

$$\mathbb{P}[(\psi_{m+1} - \bar{\psi}_m + \delta)^2 = c^2 S_m^2]$$

$$\begin{aligned}
&= \mathbb{P} \left[\boldsymbol{\xi}^\top \mathbf{D} \mathbf{V} \mathbf{D} \boldsymbol{\xi} - 2m\delta \mathbf{a}^\top \mathbf{D} \boldsymbol{\xi} - m\delta^2 = 0 \right] = \mathbb{P} \left[\boldsymbol{\zeta}^\top \boldsymbol{\Lambda} \boldsymbol{\zeta} - 2m\delta \mathbf{a}^\top \mathbf{D} \boldsymbol{\Gamma} \boldsymbol{\zeta} - m\delta^2 = 0 \right] \\
&= \mathbb{P} \left[\sum_{i=1}^{m+1} \lambda_i \zeta_i^2 - 2m\delta \mathbf{a}^\top \mathbf{D} \boldsymbol{\Gamma} \boldsymbol{\zeta} - m\delta^2 = 0 \right]. \tag{S74}
\end{aligned}$$

Because $\sigma_{m+1} = 0$, the characteristic polynomial $f(\lambda)$ simplifies to

$$f(\lambda) = -\lambda \prod_{i=1}^m (\kappa \sigma_i^2 - \lambda) + \frac{\kappa+1}{m} \lambda \cdot \sum_{i=1}^m \left[\sigma_i^2 \prod_{j \neq i, 1 \leq j \leq m} (\kappa \sigma_j^2 - \lambda) \right],$$

where $\kappa = \frac{mc^2}{m-1}$. Let $\sigma_{(1)} \leq \sigma_{(2)} \leq \dots \leq \sigma_{(m)}$ be the sorted values of $\{\sigma_i\}_{i=1}^m$. We then have

$$\begin{aligned}
(-1)^{m-1} f(\kappa \sigma_{(m)}^2) &= \frac{\kappa+1}{m} \lambda \cdot \sigma_{(m)}^2 \kappa^{m-1} \prod_{j \neq m, 1 \leq j \leq m} \{\sigma_{(m)}^2 - \sigma_{(j)}^2\} \geq 0, \\
(-1)^{m-1} f(\kappa \sigma_{(m-1)}^2) &= \frac{\kappa+1}{m} \lambda \cdot \sigma_{(m-1)}^2 \kappa^{m-1} \prod_{j \neq m-1, 1 \leq j \leq m} \{\sigma_{(m-1)}^2 - \sigma_{(j)}^2\} \leq 0,
\end{aligned}$$

which imply that $f(\lambda)$ must have a root in $[\kappa \sigma_{(m-1)}^2, \kappa \sigma_{(m)}^2]$. If $\sigma_{(m-1)}^2 > 0$, then $f(\lambda)$ must have a positive root. By a conditioning argument similar to (S73), we can know that the quantity in (S74) must be zero. If $\sigma_{(m-1)}^2 = 0$, then we have must $\sigma_{(j)}^2 = 0$ for all $1 \leq j < m$, under which $f(\lambda)$ further simplifies to

$$\begin{aligned}
f(\lambda) &= (-\lambda)^m \cdot (\kappa \sigma_{(m)}^2 - \lambda) + \frac{\kappa+1}{m} \lambda \cdot \sigma_{(m)}^2 (-\lambda)^{m-1} \\
&= (-\lambda)^m \left(\kappa \sigma_{(m)}^2 - \lambda - \frac{\kappa+1}{m} \sigma_{(m)}^2 \right) \\
&= (-\lambda)^m \frac{1}{m} \left\{ [(m-1)\kappa - 1] \sigma_{(m)}^2 - m\lambda \right\} \\
&= (-\lambda)^m \frac{1}{m} \left[(mc^2 - 1) \sigma_{(m)}^2 - m\lambda \right],
\end{aligned}$$

where the last equality follows from the definition of κ . Because $c \neq m^{-1/2}$ and $\{\sigma_j^2\}_{j=1}^{m+1}$ are not all zero, we can know that $\frac{(mc^2-1)\sigma_{(m)}^2}{m}$ must be a nonzero root of $f(\lambda)$. By a conditioning argument similar to (S73), the quantity in (S74) must be zero.

From the above, Lemma S2.1 holds. \square

Proof of Theorem 4.1. Let $\tilde{\theta}_{n,j} \equiv \sqrt{n}(\hat{\theta}_j - \mu_0)$ for $1 \leq j \leq m$, $\tilde{\theta}_{n,m+1} \equiv \sqrt{n}(\hat{\theta}_{m+1} - \mu_1)$,

and $\delta_n \equiv \sqrt{n}(\mu_1 - \mu_0)$. By definition, we can verify that

$$|\widehat{T}_m| > c \iff (\widehat{\theta}_{m+1} - \widehat{\theta}_m)^2 - c^2 \widehat{S}_m^2 > 0 \iff (\tilde{\theta}_{n,m+1} - \bar{\theta}_{n,m} + \delta_n)^2 - c^2 \tilde{S}_{n,m}^2 > 0,$$

where $\bar{\theta}_{n,m} = m^{-1} \sum_{j=1}^m \tilde{\theta}_{n,j}$ and $\tilde{S}_{n,m}^2$ is the sample variance of $\{\tilde{\theta}_{n,j}\}_{j=1}^m$. From Assumption 3.1, the condition that $\sqrt{n}(\mu_1 - \mu_0) \rightarrow \delta$ as $n \rightarrow \infty$, and using continuous mapping theorem, we have

$$(\tilde{\theta}_{n,m+1} - \bar{\theta}_{n,m} + \delta_n)^2 - c^2 \tilde{S}_{n,m}^2 \xrightarrow{d} (\psi_{m+1} - \bar{\psi}_m + \delta)^2 - c^2 S_m^2,$$

where $\psi_i \sim \mathcal{N}(0, \sigma_i^2)$ for $1 \leq j \leq m+1$, $\{\psi_i\}_{i=1}^{m+1}$ are mutually independent, and $\bar{\psi}_m$ and S_m^2 are the sample average and sample variance of $\{\psi_i\}_{i=1}^m$. From Lemma S2.1, when $c \neq m^{-1/2}$, the distribution function of $(\psi_{m+1} - \bar{\psi}_m + \delta)^2 - c^2 S_m^2$ is continuous at 0. Consequently, we must have

$$\begin{aligned} \mathbb{P}[|\widehat{T}_m| > c] &= \mathbb{P}\left[(\tilde{\theta}_{n,m+1} - \bar{\theta}_{n,m} + \delta_n)^2 - c^2 \tilde{S}_{n,m}^2 > 0\right] \\ &\rightarrow \mathbb{P}\left[(\psi_{m+1} - \bar{\psi}_m + \delta)^2 - c^2 S_m^2 > 0\right] = \mathbb{P}[|T_m| > c], \end{aligned}$$

where the last equality holds by definition. From the above, we derive Theorem 4.1. \square

S2.2 Proof of Lemmas 4.1 and A.2 and Theorem 4.3

Proof of Lemma 4.1. Lemma 4.1 follows directly from Lemmas S1.4, S1.5, S1.12 and S1.13. \square

To prove Lemma A.1, we need the following lemma.

Lemma S2.2. Let $\{\xi_i\}_{i=1}^{m+1}$ be i.i.d. standard normal random variables. For any $c > 0$ and $(\sigma_{m+1}, \sigma_1, \dots, \sigma_m) \in \mathbb{R}_{\geq 0}^{m+1}$, define

$$\begin{aligned} D_c(\sigma_{m+1}, \sigma_1, \dots, \sigma_m) &\equiv \left(\sigma_{m+1} \xi_{m+1} - m^{-1} \sum_{i=1}^m \sigma_i \xi_i \right)^2 \\ &\quad - c^2 \cdot (m-1)^{-1} \sum_{j=1}^m \left(\sigma_j \xi_j - m^{-1} \sum_{i=1}^m \sigma_i \xi_i \right)^2, \\ p_c(\sigma_{m+1}, \sigma_1, \dots, \sigma_m) &\equiv \mathbb{P}[D_c(\sigma_{m+1}, \sigma_1, \dots, \sigma_m) > 0]. \end{aligned}$$

Consider any given $c > 0, \rho > 0$ and $1 \leq k \leq m$, and let \tilde{p} denote the supremum of $p_c(\sigma_{m+1}, \sigma_1, \dots, \sigma_m)$ over all possible values of $(\sigma_{m+1}, \sigma_1, \dots, \sigma_m) \in \mathbb{R}_{\geq 0}^{m+1}$ such that $\sigma_{m+1} \leq$

$\rho\sigma_{(k)}$, where $\sigma_{(1)} \leq \sigma_{(2)} \leq \dots \leq \sigma_{(m)}$ denote the sorted values of $\{\sigma_i\}_{i=1}^m$. That is,

$$\tilde{p} \equiv \sup_{(\sigma_{m+1}, \sigma_1, \dots, \sigma_m) \in \mathbb{R}_{\geq 0}^{m+1} : \sigma_{m+1} \leq \rho\sigma_{(k)}} p_c(\sigma_{m+1}, \sigma_1, \dots, \sigma_m).$$

If $c \neq m^{-1/2}$, then one of the following two must hold:

- (a) $\tilde{p} = \sup_{(\sigma_1, \dots, \sigma_m) \in \mathbb{R}_{\geq 0}^m} p_c(0, \sigma_1, \dots, \sigma_m)$,
- (b) $\tilde{p} = p_c(\tilde{\sigma}_{m+1}, \tilde{\sigma}_1, \dots, \tilde{\sigma}_m)$ for some $(\tilde{\sigma}_{m+1}, \tilde{\sigma}_1, \dots, \tilde{\sigma}_m) \in \mathbb{R}_+ \times \mathbb{R}_{\geq 0}^m$ such that $\tilde{\sigma}_{m+1} \leq \rho\tilde{\sigma}_{(k)}$, where $\tilde{\sigma}_{(1)} \leq \tilde{\sigma}_{(2)} \leq \dots \leq \tilde{\sigma}_{(m)}$ denote the sorted values of $\{\tilde{\sigma}_i\}_{i=1}^m$.

Proof of Lemma S2.2. Below we consider any given $c > 0, \rho > 0$ and $1 \leq k \leq m$. Below we state two properties about \tilde{p} , followed from its definition.

- For any $(\sigma_{m+1}, \sigma_1, \dots, \sigma_m) \in \mathbb{R}_{\geq 0}^{m+1}$, if $\sigma_{m+1} = 0$, then we must have $\sigma_{m+1} \leq \rho\sigma_{(k)}$. By the definition of \tilde{p} , this implies that $\tilde{p} \geq \sup_{(\sigma_1, \dots, \sigma_m) \in \mathbb{R}_{\geq 0}^m} p_c(0, \sigma_1, \dots, \sigma_m)$.
- By the definition of \tilde{p} and note that the value of $p_c(\sigma_{m+1}, \sigma_1, \dots, \sigma_m)$ is invariant under permutations of $(\sigma_1, \dots, \sigma_m)$, there exists a sequence $\{(\sigma_{n,m+1}, \sigma_{n1}, \dots, \sigma_{nm})\}_{n=1}^\infty$ such that $p_n \equiv p_c(\sigma_{n,m+1}, \sigma_{n1}, \dots, \sigma_{nm}) \rightarrow \tilde{p}$ as $n \rightarrow \infty$ and $\sigma_{n,m+1} \leq \rho\sigma_{ni}$ for all n and all $k \leq i \leq n$.

First, we consider the case where there are infinitely many n such that $\sigma_{n,m+1}^2 = 0$. In this case, we must have $\tilde{p} \leq \sup_{(\sigma_1, \dots, \sigma_m) \in \mathbb{R}_{\geq 0}^m} p_c(0, \sigma_1, \dots, \sigma_m)$. From the discussion before, we must have $\tilde{p} = \sup_{(\sigma_1, \dots, \sigma_m) \in \mathbb{R}_{\geq 0}^m} p_c(0, \sigma_1, \dots, \sigma_m)$, i.e., (a) in Lemma S2.2 holds.

Second, we consider the case where there are only finitely many n such that $\sigma_{n,m+1}^2 = 0$. We can therefore assume $\sigma_{n,m+1}^2 > 0$ for all n without losing any generality. In the following, we consider two cases, depending on whether the limit superior of $\max_{1 \leq i \leq m} \frac{\sigma_{ni}}{\sigma_{n,m+1}}$ is finite.

- (i) We consider the case where the limit superior of $\max_{1 \leq i \leq m} \frac{\sigma_{ni}}{\sigma_{n,m+1}}$ is finite. By the Bolzano–Weierstrass theorem, there exists a subsequence $\{(\sigma_{n_j, m+1}, \sigma_{n_j 1}, \dots, \sigma_{n_j m})\}_{j=1}^\infty$ such that $\frac{\sigma_{ni}}{\sigma_{n,m+1}}$ converges to some a_i for each $1 \leq i \leq m$. This implies that, as $j \rightarrow \infty$,

$$D_c(1, \frac{\sigma_{n_j 1}}{\sigma_{n_j, m+1}}, \dots, \frac{\sigma_{n_j m}}{\sigma_{n_j, m+1}}) \xrightarrow{\text{a.s.}} D_c(1, a_1, \dots, a_m).$$

Because $c \neq m^{-1/2}$, Lemma S2.1 implies that

the distribution function of $D_c(1, a_1, \dots, a_m)$ is continuous at zero. By the property

of weak convergence, we then have

$$\begin{aligned}\tilde{p} &= \lim_{j \rightarrow \infty} p_{n_j} = \lim_{j \rightarrow \infty} \mathbb{P}[D_c(1, \frac{\sigma_{n_j 1}}{\sigma_{n_j, m+1}}, \dots, \frac{\sigma_{n_j m}}{\sigma_{n_j, m+1}}) > 0] = \mathbb{P}[D_c(1, a_1, \dots, a_n) > 0] \\ &= p_c(1, a_1, \dots, a_n).\end{aligned}$$

Furthermore, it is easy to verify that $\rho a_i = \rho \lim_{j \rightarrow \infty} \frac{\sigma_{n_j i}}{\sigma_{n_j, m+1}} \geq 1$ for $k \leq i \leq n$. Consequently, (b) in Lemma S2.2 holds.

- (ii) We consider the case where the limit superior of $\max_{1 \leq i \leq m} \frac{\sigma_{ni}}{\sigma_{n, m+1}}$ is infinite. There then exists a subsequence such that $\frac{\sigma_{n, m+1}}{\max_{1 \leq i \leq m} \sigma_{ni}} \rightarrow 0$ along this subsequence. By the Bolzano–Weierstrass theorem, there exists a subsequence $\{(\sigma_{n_j, m+1}, \sigma_{n_j 1}, \dots, \sigma_{n_j m})\}_{j=1}^{\infty}$ such that, as $j \rightarrow \infty$, $\frac{\sigma_{n_j, m+1}}{\max_{1 \leq l \leq m} \sigma_{n_j l}} \rightarrow 0$ and $\frac{\sigma_{n_j i}}{\max_{1 \leq l \leq m} \sigma_{n_j l}} \rightarrow b_i \leq 1$ for all $1 \leq i \leq m$. Moreover, at least one of $\{b_i\}_{i=1}^m$ is 1.

These then imply that

$$D_c(\frac{\sigma_{n_j, m+1}}{\max_{1 \leq l \leq m} \sigma_{n_j l}}, \frac{\sigma_{n_j 1}}{\max_{1 \leq l \leq m} \sigma_{n_j l}}, \dots, \frac{\sigma_{n_j m}}{\max_{1 \leq l \leq m} \sigma_{n_j l}}) \xrightarrow{\text{a.s.}} D_c(0, b_1, \dots, b_n).$$

Since $c \neq m^{-1/2}$ holds, Lemma S2.1 implies that the distribution function of $D_c(0, b_1, \dots, b_n)$ is continuous at zero. By the property of weak convergence, we then have

$$\begin{aligned}\tilde{p} &= \lim_{j \rightarrow \infty} p_{n_j} = \lim_{j \rightarrow \infty} \mathbb{P}[D_c(\frac{\sigma_{n_j, m+1}}{\max_{1 \leq l \leq m} \sigma_{n_j l}}, \frac{\sigma_{n_j 1}}{\max_{1 \leq l \leq m} \sigma_{n_j l}}, \dots, \frac{\sigma_{n_j m}}{\max_{1 \leq l \leq m} \sigma_{n_j l}}) > 0] \\ &= \mathbb{P}[D_c(0, b_1, \dots, b_n) > 0] = p_c(0, b_1, \dots, b_n) \leq \sup_{(\sigma_1, \dots, \sigma_m) \in \mathbb{R}_{\geq 0}^m} p_c(0, \sigma_1, \dots, \sigma_m).\end{aligned}$$

From the discussion before, this implies that $\tilde{p} = \sup_{(\sigma_1, \dots, \sigma_m) \in \mathbb{R}_{\geq 0}^m} p_c(0, \sigma_1, \dots, \sigma_m)$, i.e., (a) in Lemma S2.2 holds.

From the above, Lemma S2.2 holds. □

Lemma S2.3. For any $1 \leq k \leq m$ and $c \neq m^{-1/2}$,¹² the maximum rejection probability $p_m(c; k, \rho)$ in (11) must be obtained at some $(\sigma_1, \dots, \sigma_m, \sigma_{m+1}) \in \mathcal{S}_m(k, \rho)$.

Proof of Lemma A.1. From Lemma S2.2, we can know that one of the following two must hold:

¹²Similar to the footnote for Theorem 4.1 and as discussed in Remark 4.4, we will consider values of c greater than $m^{-1/2}$ for most conventional significance levels.

- (a) $p_m(c; k, \rho) = \sup_{(\sigma_1, \dots, \sigma_m) \in \mathbb{R}_{\geq 0}^m} p_c(0, \sigma_1, \dots, \sigma_m)$,
- (b) $p_m(c; k, \rho) = p_c(\tilde{\sigma}_{m+1}, \tilde{\sigma}_1, \dots, \tilde{\sigma}_m)$ for some $(\tilde{\sigma}_{m+1}, \tilde{\sigma}_1, \dots, \tilde{\sigma}_m) \in \mathbb{R}_+ \times \mathbb{R}_{\geq 0}^m$ such that $\tilde{\sigma}_{m+1} \leq \rho \tilde{\sigma}_{(k)}$, where $\tilde{\sigma}_{(1)} \leq \tilde{\sigma}_{(2)} \leq \dots \leq \tilde{\sigma}_{(m)}$ denote the sorted values of $\{\tilde{\sigma}_i\}_{i=1}^m$.

When (b) holds, Lemma A.1 holds obviously. Below we consider only the case when (a) holds. When $\sigma_{m+1} = 0$, our t statistic essentially reduces to a one-sample t statistic, except for a constant scaling term of \sqrt{m} . From Bakirov and Székely (2006), the supremum of $p_c(0, \sigma_1, \dots, \sigma_m)$ over $(\sigma_1, \dots, \sigma_m) \in \mathbb{R}_{\geq 0}^m$ must be achieved when some of $\{\sigma_j\}_{j=1}^m$ are zero and the remaining take a some common positive value, such as 1. From the above, Lemma A.1 holds. \square

Proof of Lemma A.2. Lemma A.2 follows directly from Lemma S1.11, noting that $z = \kappa \gamma_1^2$, where κ is a constant depending only on m and c . \square

Proof of Theorem 4.2. Consider any given $1 \leq k \leq m$, $\rho \geq 0$, and $c > 0$ with $c \neq m^{-1/2}$. When $\rho = 0$, Theorem 4.2 follows immediately from Bakirov and Székely (2006). We then consider the case where $\rho > 0$. From Lemma A.1, the maximum rejection probability $p_m(c; k, \rho)$ is achieved at some finite $\{\sigma_j\}_{j=1}^m$. If $\sigma_{m+1} = 0$, then the maximizer must have the form in (i), as shown in Bakirov and Székely (2006). Below we consider only the case where $\sigma_{m+1} > 0$.

Because the rejection probability is unchanged when we scale all the variances $\{\sigma_j\}_{j=1}^m$ by a positive constant, we must have $p_m(c; k, \rho) = \bar{p}_m(c; \gamma_1, \dots, \gamma_m)$ for some $(\gamma_1, \dots, \gamma_m) \in \mathbb{R}_{\geq 0}^m$, where $\gamma_{(k)} \geq \rho^{-1}$ and $\gamma_{(1)} \leq \gamma_{(2)} \leq \dots \leq \gamma_{(m)}$ are the sorted values of $\{\gamma_i\}_{i=1}^m$. We prove that, besides 0 and ρ^{-1} , $\{\gamma_i\}_{i=1}^m$ cannot take more than one distinct values; equivalently, there exists $\gamma \geq 0$ such that $\gamma_i \in \{0, \rho^{-1}, \gamma\}$ for all $1 \leq i \leq m$. We prove this by contradiction. Assume that, without loss of generality, $\gamma_1, \gamma_2 \notin \{0, \rho^{-1}\}$, and $\gamma_1 \neq \gamma_2$. Fix the values of $c, \gamma_3, \dots, \gamma_m$ and θ_{m+1} , and view $\bar{p}_m(c; \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_m)$ as a function of only γ_1 , where γ_2 is uniquely determined by $\gamma_1, c, \gamma_3, \dots, \gamma_m$ and θ_{m+1} . Note that the condition $\gamma_{(k)} \geq \rho^{-1}$ still holds when we slightly change the values of γ_1 and consequently γ_2 . Thus, γ_1 is at least a local maximizer of $\bar{p}_m(c; \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_m)$ over a sufficiently small neighborhood of γ_1 . This implies that the first order derivative of $\bar{p}_m(c; \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_m)$ over γ_1^2 is zero, and the second order derivative of $\bar{p}_m(c; \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_m)$ over γ_1^2 is less than or equal to zero. This, however, contradicts Lemma A.2.

Therefore, we must have $p_m(c; k, \rho) = \bar{p}_m(c; \gamma_1, \dots, \gamma_m)$ for some $\{\gamma_j\}_{j=1}^m$ such that $\gamma_i \in \{0, \rho^{-1}, \gamma\}$ for $1 \leq i \leq m$ and some $\gamma \geq 0$. Again, because the rejection probability is unchanged when we scale all the variances $\{\sigma_j\}_{j=1}^m$ by a positive constant, the maximum

rejection probability must be obtained at some $\{\sigma_j\}_{j=1}^{m+1}$ such that $\sigma_{m+1} = \rho$, and $\sigma_i \in \{0, 1, \gamma\}$ for $1 \leq i \leq m$ and some $\gamma \geq 0$.

From the above, Theorem 4.2 holds. \square

Proof of Theorem 4.3. Consider any given $1 \leq k \leq m$, $\rho > 0$, and $c > 0$ with $c \neq m^{-1/2}$. From Theorem 4.2, we have either $p_m(c; k, \rho) = p_{m,0}(c)$, or $p_m(c; k, \rho) = \bar{p}_m(c; \gamma_1, \dots, \gamma_m)$ for some $(\gamma_1, \dots, \gamma_m) \in \mathbb{R}_{\geq 0}^m$ such that $\gamma_i \in \{0, \rho^{-1}, \gamma\}$ for some $\gamma \in \mathbb{R}$ and $\gamma_{(k)} \geq \rho^{-1}$, where $\gamma_{(1)} \leq \gamma_{(2)} \leq \dots \leq \gamma_{(m)}$ are the sorted values of $\{\gamma_i\}_{i=1}^m$.

Now suppose that the latter holds, i.e., $p_m(c; k, \rho) = \bar{p}_m(c; \gamma_1, \dots, \gamma_m)$ some $(\gamma_1, \dots, \gamma_m) \in \mathbb{R}_{\geq 0}^m$ satisfying the conditions discussed before.

Let m_1 and m_0 denote the numbers of γ_i s that take values ρ^{-1} and 0, respectively. That is, $m_1 = \sum_{i=1}^m \mathbf{1}(\gamma_i = \rho^{-1})$ and $m_0 = \sum_{i=1}^m \mathbf{1}(\gamma_i = 0)$. Consequently, the number of γ_i s that take the value γ is $m - m_1 - m_0$, and $\bar{p}_m(c; \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_m)$ simplifies to $\bar{p}_m(c; \rho, \gamma; m_1, m_0)$ defined in (15). Note that $\{\gamma_i\}_{i=1}^m$ needs to satisfy the constraint that $\gamma_{(k)} \geq \rho^{-1}$. Thus, we must have $0 \leq m_0 \leq k - 1$, $m_1 \leq m - m_0$, and $\gamma \in \mathbb{R}_{\geq 0}$ if $m_1 \geq m - k + 1$ and $\gamma \in [\rho^{-1}, \infty)$ if $m_1 < m - k + 1$. In sum, $p_m(c; k, \rho)$ equals to $\bar{p}_m(c; \rho, \gamma; m_1, m_0)$ for some $0 \leq m_0 \leq k - 1$, $m_1 \leq m - m_0$, and $\gamma \in \mathbb{R}_{\geq 0}$ if $m_1 \geq m - k + 1$ and $\gamma \in [\rho^{-1}, \infty)$ if $m_1 < m - k + 1$.

From the above, we list possible cases where the rejection probability obtains its supremum value $p_m(c; k, \rho)$. Thus, $p_m(c; k, \rho)$ must be the supremum over all the cases we discussed above. We can therefore derive Theorem 4.3. \square

S2.3 Proof of Lemma A.3

Below we first give the form of $H_m(c; \gamma_1, \dots, \gamma_m)$ in Lemma A.3. For any $c > 0$ and $(\gamma_1, \gamma_2, \dots, \gamma_m) \in \mathbb{R}_{\geq 0}^m$, let $\kappa \equiv \frac{mc^2}{m-1}$, $\tau \equiv \frac{\kappa+1}{\kappa m}$, $x_i \equiv \kappa \gamma_i^2$ for $1 \leq i \leq m$, $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(m)}$ be the sorted values of $\{x_i\}_{i=1}^m$, and

$$\begin{aligned} & H'_m(w; c; \gamma_1, \dots, \gamma_m) \\ & \equiv \frac{\bar{\theta}}{x_{(m)} + \bar{\theta}} + \frac{2\bar{\theta}}{w + \bar{\theta}} - \frac{\bar{\theta} - m}{\tau\bar{\theta} - 1} + \frac{1 - \tau}{1 - \tau + \min\{C_m(w; c; \gamma_1, \dots, \gamma_m), 0\}} - 1, \end{aligned} \quad (\text{S75})$$

where $\bar{\theta} \equiv m + \frac{x_{(m)}}{\kappa}$, $\underline{\theta} \equiv \max_{1 \leq j \leq m} \theta(x_{(j)}, j)$ with $\theta(\cdot)$ defined as in Lemma S1.12, and

$$C_m(w; c; \gamma_1, \dots, \gamma_m) \equiv \begin{cases} \left(1 - 2\tau + \frac{2}{x_{(m)}}\right) \frac{wx_{(m)} - \bar{\theta}^2}{(w + x_{(m)} + 2\bar{\theta})\bar{\theta}} + \frac{1}{x_{(m)}} - \tau, & \text{if } w = x_{(1)}, \\ (1 - 2\tau) \frac{wz - \bar{\theta}^2}{(w + z + 2\bar{\theta})\bar{\theta}} - \tau, & \text{if } w > x_{(1)}. \end{cases}$$

Then $H_m(c; \gamma_1, \dots, \gamma_m)$ is defined as

$$H_m(c; \gamma_1, \dots, \gamma_m) = H'_m(x_2; c; \gamma_1, \dots, \gamma_m). \quad (\text{S76})$$

Proof of Lemma A.3. Note that if $c \geq \sqrt{2(m-1)/[m(m-2)]}$, then τ defined as in Lemma S1.19 satisfies that

$$\tau = \frac{\kappa + 1}{\kappa m} = \frac{1}{m} + \frac{1}{\kappa m} = \frac{1}{m} + \frac{m-1}{m^2 c^2} \leq \frac{1}{m} + \frac{(m-1)m(m-2)}{m^2 \cdot 2(m-1)} = \frac{1}{m} + \frac{m-2}{2m} = \frac{1}{2},$$

where κ is defined as in Lemma S1.19. Lemma A.3 follows immediately from Lemma S1.19. \square

S2.4 Simplifying the optimization under general relative heterogeneity assumption

Below we consider simplification on optimization under general relative heterogeneity constraint. Specifically, Lemma A.3 can help to simplify the optimization in (16). Consider any $1 \leq m_1 \leq m-1$, $0 \leq m_0 \leq m-m_1-1$, and the optimization of $\bar{p}_m(c; \rho, \gamma; m_1, m_0)$ over either $\gamma \in [0, \infty)$ or $\gamma \in [\rho^{-1}, \infty)$. Let

$$(\gamma_1, \gamma_2, \dots, \gamma_m) = (\underbrace{\gamma, \dots, \gamma}_{m-m_1-m_0}, \underbrace{\rho^{-1}, \dots, \rho^{-1}}_{m_1}, \underbrace{0, \dots, 0}_{m_0}).$$

Define

$$\check{H}_m(c; \rho, \gamma; m_1, m_0) = H'_m(\rho^{-1}; c; \gamma_1, \dots, \gamma_m)\}$$

For any $\rho > 0$ and $\gamma > \rho^{-1}$, if $\check{H}_m(c; \rho, \gamma; m_1, m_0) < 0$, then we can strictly increase the rejection probability $\bar{p}_m(\gamma_1, \gamma_2, \dots, \gamma_m)$ by slightly decreasing one of $\{\gamma_j\}_{j=1}^m$ that is equal to γ and slightly increasing one of $\{\gamma_j\}_{j=1}^m$ that is equal to ρ^{-1} . Importantly, if the original $\{\gamma_j\}_{j=1}^m$ satisfy the relative heterogeneity assumption for some k , then the slight changes of the $\{\gamma_j\}_{j=1}^m$ will maintain this relative heterogeneity assumption. Con-

sequently, $\bar{p}_m(\gamma_1, \gamma_2, \dots, \gamma_m)$ cannot be the maximum rejection probability under the relative heterogeneity assumption.

From the above, if $\check{H}_m(c; \rho, \gamma; m_1, m_0) < 0$ for all $\gamma > \rho^{-1}$, then $\sup_{\gamma \in [\rho^{-1}, \infty)} \bar{p}_m(c; \rho, \gamma; m_1, m_0)$ must be obtained at $\gamma = \rho^{-1}$, and $\sup_{\gamma \in \mathbb{R}_{\geq 0}} \bar{p}_m(c; \rho, \gamma; m_1, m_0)$ must be obtained at some $\gamma \in [0, \rho^{-1}]$. In other words, we can either obtain a closed-form solution for the optimization or restrict the optimization to a smaller range.

S2.5 Simplifying the optimization under relative heterogeneity assumption with $k = 1$

Theorem S1. For any given $m \geq 4$, $\rho > 0$, $c > \sqrt{\frac{3(m-1)}{m(m-3)}}$, define the following as a function of γ :

$$\tilde{H}_m(\gamma; c, \rho) \equiv \frac{\bar{\theta}}{\kappa\gamma^2 + \bar{\theta}} + \frac{2\bar{\theta}}{\kappa\rho^{-2} + \bar{\theta}} - \frac{\underline{\theta} - m}{\tau\underline{\theta} - 1} + \frac{1 - \tau}{1 - \tau + \min\{C, 0\}} - 1,$$

where $\bar{\theta} \equiv m + \gamma^2$, $\underline{\theta} \equiv m + \rho^{-2}$, $\kappa \equiv \frac{mc^2}{m-1}$, $\tau \equiv \frac{\kappa+1}{m\kappa}$, and

$$C \equiv \left(1 - 2\tau + \frac{2}{\kappa\gamma^2}\right) \frac{\kappa^2\gamma^2\rho^{-2} - \bar{\theta}^2}{(\kappa\gamma^2 + \kappa\rho^{-2} + 2\bar{\theta})\bar{\theta}} + \frac{1}{\kappa\gamma^2} - \tau.$$

Suppose that

$$\tilde{H}_m(\gamma; c, \rho) < 0 \text{ for all } \gamma > \rho^{-1}.$$

Then, the maximum rejection probability $p_m(c; 1, \rho)$ under Assumption 3.2 with $k = 1$ and the given ρ , m , and c has the following equivalent form:

$$p_m(c; 1, \rho) = \mathbb{P} \left[|t_{m-1}| \sqrt{\rho^2 + \frac{1}{m}} > c \right].$$

Proof of Theorem S1. First, for any given $1 \leq m_1 \leq m - 1$, $\rho > 0$ and $\gamma > \rho^{-1}$, we consider

$$(\gamma_1, \gamma_2, \dots, \gamma_m) = (\gamma, \rho^{-1}, \underbrace{\gamma, \dots, \gamma}_{m-m_1-1}, \underbrace{\rho^{-1}, \dots, \rho^{-1}}_{m_1-1}). \quad (\text{S77})$$

Recall the definition in (S75) and (S76), we have

$$H_m(c; \gamma_1, \dots, \gamma_m)$$

$$\begin{aligned}
&= H'_m(x_2; c; \gamma_1, \dots, \gamma_m) \\
&= \frac{\bar{\theta}}{x_{(m)} + \bar{\theta}} + \frac{2\bar{\theta}}{x_2 + \bar{\theta}} - \frac{\underline{\theta}' - m}{\tau \underline{\theta}' - 1} + \frac{1 - \tau}{1 - \tau + \min\{C_m(x_2; c; \gamma_1, \dots, \gamma_m), 0\}} - 1,
\end{aligned}$$

where $\kappa \equiv \frac{mc^2}{m-1}$, $\tau \equiv \frac{\kappa+1}{\kappa m}$, $x_i \equiv \kappa \gamma_i^2$ for $1 \leq i \leq m$, $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(m)}$ are the sorted values of $\{x_i\}_{i=1}^m$, $\bar{\theta} = m + x_{(m)}/\kappa$, $\underline{\theta}' = \max_{1 \leq j \leq m} \theta(x_{(j)}, j)$ with $\theta(\cdot)$ defined as in Lemma S1.12, and

$$C_m(x_2; c; \gamma_1, \dots, \gamma_m) = \left(1 - 2\tau + \frac{2}{x_{(m)}}\right) \frac{x_2 x_{(m)} - \bar{\theta}^2}{(x_2 + x_{(m)} + 2\bar{\theta})\bar{\theta}} + \frac{1}{x_{(m)}} - \tau.$$

Note that $(\theta - m)/(\tau\theta - 1)$ is increasing in θ , due to the fact that $-1 + m\tau = 1/\kappa > 0$, and $\underline{\theta}' = \max_{1 \leq j \leq m} \theta(x_{(j)}, j) \geq \theta(x_{(1)}, 1) = m + \rho^{-2} = \underline{\theta}$, where the last equality follows from Lemma S1.12(c). We then have

$$\begin{aligned}
H_m(c; \gamma_1, \dots, \gamma_m) &= \frac{\bar{\theta}}{x_{(m)} + \bar{\theta}} + \frac{2\bar{\theta}}{x_2 + \bar{\theta}} - \frac{\underline{\theta}' - m}{\tau \underline{\theta}' - 1} + \frac{1 - \tau}{1 - \tau + \min\{C_m(x_2; c; \gamma_1, \dots, \gamma_m), 0\}} - 1 \\
&\leq \frac{\bar{\theta}}{x_{(m)} + \bar{\theta}} + \frac{2\bar{\theta}}{x_2 + \bar{\theta}} - \frac{\underline{\theta} - m}{\tau \underline{\theta} - 1} + \frac{1 - \tau}{1 - \tau + \min\{C_m(x_2; c; \gamma_1, \dots, \gamma_m), 0\}} - 1 \\
&= \tilde{H}_m(c; \rho, \gamma),
\end{aligned}$$

where the last equality follows by definition.

Second, from the first part and the condition that $\tilde{H}_m(c; \rho, \gamma) < 0$ for all $\gamma > \rho^{-1}$, we know that $H_m(c; \gamma_1, \dots, \gamma_m) < 0$, for all $1 \leq m_1 \leq m$, $\rho > 0$, $\gamma > \rho^{-1}$ and $(\gamma_1, \dots, \gamma_m)$ defined as in (S77). In addition, we have

$$\frac{c}{\sqrt{\frac{2(m-1)}{m(m-2)}}} > \frac{\sqrt{\frac{3(m-1)}{m(m-3)}}}{\sqrt{\frac{2(m-1)}{m(m-2)}}} = \sqrt{\frac{3(m-2)}{2(m-3)}} = \sqrt{\frac{3m-6}{2m-6}} \geq 1.$$

From Lemma A.3, we know that for $1 \leq m_1 \leq m-1$ and any $c, \rho > 0$, $\bar{p}_m(c; \rho, \gamma; m_1, 0)$ cannot be $p_m(c; 1, \rho)$ for any $\gamma > \rho^{-1}$, since we can strictly increase the rejection probability $\bar{p}_m(\gamma_1, \dots, \gamma_m)$ by slightly decreasing one of $\{\gamma_i\}_{i=1}^m$ that is equal to γ and slightly increasing one of $\{\gamma_i\}_{i=1}^m$ that is equal to ρ^{-1} .

Third, we consider $\bar{p}_m(c; \gamma \mathbf{1}_m^\top)$ for any $\gamma > 0$. In this case, $\bar{p}_m(c; \gamma \mathbf{1}_m^\top) = \mathbb{P}[|T_m| > c]$, where T_m is defined as in (8) with $\psi_{m+1} \sim \mathcal{N}(0, \gamma^{-2})$ and $\psi_i \sim \mathcal{N}(0, 1)$ for $1 \leq i \leq m$. We can verify that $T_m \sim \sqrt{\gamma^{-2} + m^{-1}} t_{m-1}$, where t_{m-1} is a t -distributed random variable

with $m - 1$ degrees of freedom. Consequently,

$$\bar{p}_m(c; \gamma \mathbf{1}_m^\top) = \mathbb{P}[|T_m| > c] = \mathbb{P}\left[|t_{m-1}| > \frac{c}{\sqrt{\gamma^{-2} + m^{-1}}}\right],$$

which is decreasing in γ . This implies that $\sup_{\gamma \in [\rho^{-1}, \infty)} \bar{p}_m(c; \gamma \mathbf{1}_m^\top) = \bar{p}_m(c; \rho^{-1} \mathbf{1}_m^\top)$.

Fourth, from the second and third parts, we know that $p_m(c; 1, \rho)$ defined as in Theorem 4.3 has the following equivalent forms:

$$\begin{aligned} p_m(c; 1, \rho) &= \max \left\{ \max_{0 \leq m_1 \leq m} \tilde{p}_m(c; k, \rho; m_1, 0), p_{m,0}(c) \right\} \\ &= \max \left\{ \bar{p}_m(c; \rho^{-1} \mathbf{1}_m^\top), p_{m,0}(c) \right\} \\ &= \max \left\{ \mathbb{P}\left[|t_{m-1}| > \frac{c}{\sqrt{\rho^2 + m^{-1}}}\right], p_{m,0}(c) \right\}. \end{aligned}$$

Fifth, from Bakirov and Székely (2006), we can know that, when $c > \sqrt{3(m-1)/[m(m-3)]}$,

$$p_{m,0}(c) = \mathbb{P}[|t_{m-1}| > \sqrt{mc}] \leq \mathbb{P}\left[|t_{m-1}| > \frac{c}{\sqrt{\rho^2 + m^{-1}}}\right].$$

This implies that

$$p_m(c; 1, \rho) = \max \left\{ \mathbb{P}\left[|t_{m-1}| > \frac{c}{\sqrt{\rho^2 + m^{-1}}}\right], p_{m,0}(c) \right\} = \mathbb{P}\left[|t_{m-1}| > \frac{c}{\sqrt{\rho^2 + m^{-1}}}\right].$$

Therefore, Theorem S1 holds. \square

S2.6 Proof of Theorem 4.4

We will prove Theorem 4.4 using Theorem S1. Define

$$\begin{aligned} \tilde{H}_{m,1}(\gamma; c, \rho) &\equiv \frac{\bar{\theta}}{\kappa\gamma^2 + \bar{\theta}} + \frac{2\bar{\theta}}{\kappa\rho^{-2} + \bar{\theta}}, \\ \tilde{H}_{m,2}(\gamma; c, \rho) &\equiv \frac{1 - \tau}{1 - \tau + \min\{C, 0\}}, \\ \tilde{H}_{m,3}(c, \rho) &\equiv -\frac{\underline{\theta} - m}{\tau\underline{\theta} - 1}, \end{aligned} \tag{S78}$$

where $\kappa, \bar{\theta}, \underline{\theta}, \tau$ and C are defined the same as in Theorem S1:

$$\begin{aligned}\kappa &\equiv \frac{mc^2}{m-1}, \quad \bar{\theta} \equiv m + \gamma^2, \quad \underline{\theta} \equiv m + \rho^{-2}, \quad \tau \equiv \frac{\kappa+1}{m\kappa}, \\ C &\equiv \left(1 - 2\tau + \frac{2}{\kappa\gamma^2}\right) \frac{\kappa^2\gamma^2\rho^{-2} - \bar{\theta}^2}{(\kappa\gamma^2 + \kappa\rho^{-2} + 2\bar{\theta})\bar{\theta}} + \frac{1}{\kappa\gamma^2} - \tau.\end{aligned}\tag{S79}$$

By definition, we can then write $\tilde{H}_m(\gamma; c, \rho)$ in Theorem S1 as

$$\tilde{H}_m(\gamma; c, \rho) = \tilde{H}_{m,1}(\gamma; c, \rho) + \tilde{H}_{m,2}(\gamma; c, \rho) + \tilde{H}_{m,3}(c, \rho) - 1.$$

In the following Lemmas S2.4–S2.7, we construct upper bounds $\bar{H}_{m,1}(c, \rho)$ and $\bar{H}_{m,2}(c, \rho)$ for $\tilde{H}_{m,1}(\gamma; c, \rho)$ and $\tilde{H}_{m,2}(\gamma; c, \rho)$, respectively, over all $\gamma \in (\rho^{-1}, \infty)$; see (S86) and (S96) for their expressions. We can thus bound $\tilde{H}_m(\gamma; c, \rho)$ over all $\gamma > \rho^{-1}$ from the above by

$$\bar{H}_{m,1}(c, \rho) + \bar{H}_{m,2}(c, \rho) + \tilde{H}_{m,3}(c, \rho) - 1.$$

We then establish the monotonicity of $\bar{H}_{m,1}(c, \rho)$, $\bar{H}_{m,2}(c, \rho)$, and $\tilde{H}_{m,3}(c, \rho)$ with respect to c in Lemmas S2.8–S2.10, respectively. Finally, we prove Theorem 4.4.

Lemma S2.4. Define $\tilde{H}_{m,1}(\gamma; c, \rho)$ as in (S78) for $\gamma > \rho^{-1}$. One of the following statements is true:

- (a) There exists at most one finite γ_0 such that $\gamma_0 > \rho^{-1}$ and $\left. \frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial \gamma} \right|_{\gamma=\gamma_0} = 0$.
- (b) $\frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial \gamma} = 0$ for any $\gamma > \rho^{-1}$.

Proof of Lemma S2.4. Note that $\tilde{H}_{m,1}(\gamma; c, \rho)$ can be written explicitly in terms of γ^2 as follows:

$$\tilde{H}_{m,1}(\gamma; c, \rho) \equiv \frac{m + \gamma^2}{(\kappa + 1)\gamma^2 + m} + 2 \left[\frac{m + \gamma^2}{\gamma^2 + (\kappa\rho^{-2} + m)} \right].\tag{S80}$$

The derivatives of the above two fractions with respect to γ^2 are, respectively,

$$\frac{\partial \left[\frac{m + \gamma^2}{(\kappa + 1)\gamma^2 + m} \right]}{\partial (\gamma^2)} = \frac{-m\kappa}{[(\kappa + 1)\gamma^2 + m]^2},\tag{S81}$$

and

$$\frac{\partial \left[\frac{m + \gamma^2}{\gamma^2 + (\kappa\rho^{-2} + m)} \right]}{\partial (\gamma^2)} = \frac{\kappa\rho^{-2}}{[\gamma^2 + (\kappa\rho^{-2} + m)]^2}.\tag{S82}$$

Combining (S81) and (S82) gives

$$\begin{aligned}
& \frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial(\gamma^2)} \\
&= \frac{\partial[\frac{m+\gamma^2}{(\kappa+1)\gamma^2+m}]}{\partial(\gamma^2)} + 2 \frac{\partial[\frac{m+\gamma^2}{\gamma^2+(\kappa\rho^{-2}+m)}]}{\partial(\gamma^2)} = \kappa \left\{ \frac{2\rho^{-2}}{[\gamma^2+(\kappa\rho^{-2}+m)]^2} - \frac{m}{[(\kappa+1)\gamma^2+m]^2} \right\} \\
&= \frac{\kappa\{2\rho^{-2}[(\kappa+1)\gamma^2+m]^2 - m[\gamma^2+(\kappa\rho^{-2}+m)]^2\}}{[\gamma^2+(\kappa\rho^{-2}+m)]^2 \cdot [(\kappa+1)\gamma^2+m]^2} \\
&= \frac{\kappa\{\sqrt{2\rho^{-2}}[(\kappa+1)\gamma^2+m] + \sqrt{m}[\gamma^2+(\kappa\rho^{-2}+m)]\}}{[\gamma^2+(\kappa\rho^{-2}+m)]^2 \cdot [(\kappa+1)\gamma^2+m]^2} \times \\
&\quad \left\{ \sqrt{2\rho^{-2}}[(\kappa+1)\gamma^2+m] - \sqrt{m}[\gamma^2+(\kappa\rho^{-2}+m)] \right\} \\
&= \frac{\kappa\{\sqrt{2\rho^{-2}}[(\kappa+1)\gamma^2+m] + \sqrt{m}[\gamma^2+(\kappa\rho^{-2}+m)]\}}{[\gamma^2+(\kappa\rho^{-2}+m)]^2 \cdot [(\kappa+1)\gamma^2+m]^2} \times \\
&\quad \left\{ \left[\sqrt{2\rho^{-2}}(\kappa+1) - \sqrt{m} \right] \gamma^2 - \left[\sqrt{m}(\kappa\rho^{-2}+m) - \sqrt{2\rho^{-2}}m \right] \right\}. \tag{S83}
\end{aligned}$$

The goal of this lemma is to show either there exists at most one $\gamma_0 > \rho^{-1}$ that satisfy statement (a), or the partial derivative is constant for all γ . Since $\frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial(\gamma^2)} = 2\gamma \frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial \gamma}$ and we require $\gamma > \rho^{-1} > 0$, finding a γ_0 such that $\frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial \gamma} \Big|_{\gamma=\gamma_0} = 0$ is equivalent to finding γ_0 that satisfy $\frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial(\gamma^2)} \Big|_{\gamma^2=\gamma_0^2} = 0$. To this end, it suffices to consider the scenario such that (S83) equals 0. Because ρ, κ, γ and m are all positive, (S83) equals 0 if and only if

$$\left[\sqrt{2\rho^{-2}}(\kappa+1) - \sqrt{m} \right] \gamma^2 = \sqrt{m}(\kappa\rho^{-2}+m) - \sqrt{2\rho^{-2}}m. \tag{S84}$$

There are two scenarios to consider.

First, consider the case where $\sqrt{2\rho^{-2}}(\kappa+1) - \sqrt{m} \neq 0$. Then, (S84) implies that

$$\gamma^2 = \frac{\sqrt{m}(\kappa\rho^{-2}+m) - \sqrt{2\rho^{-2}}m}{\sqrt{2\rho^{-2}}(\kappa+1) - \sqrt{m}}. \tag{S85}$$

In (S85), if the RHS is positive, then there is one positive γ that satisfy the equation. If the RHS equals 0, then this cannot be the γ_0 we wish to find because we require $\gamma_0 > \rho^{-1} > 0$. If the RHS is negative, then there is no real solution. Thus, in this case,

there is at most one $\gamma_0 > \rho^{-1}$ such that $\left. \frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial \gamma} \right|_{\gamma=\gamma_0} = 0$.

Now, consider the second scenario where $\sqrt{2\rho^{-2}}(\kappa + 1) - \sqrt{m} = 0$ under the first case. Then, the derivative is either always positive, negative, or zero for all $\gamma > \rho^{-1}$. Thus, either statement (a) or (b) is true.

From the above, Lemma S2.4 holds. \square

Lemma S2.5. For any given values of positive c , ρ , and m , define $\tilde{H}_{m,1}(\gamma; c, \rho)$ as in (S78) for $\gamma > \rho^{-1}$. For any $\gamma > \rho^{-1}$, we have $\tilde{H}_{m,1}(\gamma; c, \rho) \leq \bar{H}_{m,1}(c, \rho)$, where

$$\begin{aligned} \bar{H}_{m,1}(c, \rho) &\equiv \max \left\{ \tilde{H}_{m,1}(\rho^{-1}; c, \rho), \lim_{\gamma \rightarrow \infty} \tilde{H}_{m,1}(\gamma; c, \rho) \right\} \\ &= \max \left\{ \frac{3(m + \rho^{-2})}{(\kappa + 1)\rho^{-2} + m}, \frac{2\kappa + 3}{\kappa + 1} \right\}. \end{aligned} \quad (\text{S86})$$

Proof of Lemma S2.5. Following Lemma S2.4, we consider the following three scenarios.

First, suppose there is no $\gamma_0 > \rho^{-1}$ such that $\left. \frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial \gamma} \right|_{\gamma=\gamma_0} = 0$. This means $\frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial \gamma}$ is either positive or negative for all $\gamma > \rho^{-1}$ because $\frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial \gamma}$ is continuous for all real γ from (S83). Thus, $\tilde{H}_{m,1}(\gamma; c, \rho)$ is monotone in γ for all $\gamma > \rho^{-1}$.

Second, suppose that $\frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial \gamma} = 0$ for any $\gamma > \rho^{-1}$. This means $\tilde{H}_{m,1}(\gamma; c, \rho)$ is constant in γ for all $\gamma > \rho^{-1}$.

Third, suppose there is one $\gamma_0 > \rho^{-1}$ such that $\left. \frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial \gamma} \right|_{\gamma=\gamma_0} = 0$. For this γ_0 , it has to satisfy (S85). There are two scenarios to consider as follows:

- (i) $\sqrt{m}(\kappa\rho^{-2} + m) - \sqrt{2\rho^{-2}}m > 0$ and $\sqrt{2\rho^{-2}}(\kappa + 1) - \sqrt{m} > 0$.
- (ii) $\sqrt{m}(\kappa\rho^{-2} + m) - \sqrt{2\rho^{-2}}m < 0$ and $\sqrt{2\rho^{-2}}(\kappa + 1) - \sqrt{m} < 0$.

We first show that scenario (ii) is impossible. To see this, assume to the contrary that $\sqrt{m}(\kappa\rho^{-2} + m) - \sqrt{2\rho^{-2}}m < 0$ and $\sqrt{2\rho^{-2}}(\kappa + 1) - \sqrt{m} < 0$. The first inequality implies that

$$\kappa\rho^{-2} + m < \sqrt{2\rho^{-2}}m, \quad (\text{S87})$$

whereas the second inequality implies that

$$\sqrt{2\rho^{-2}} < \frac{\sqrt{m}}{\kappa + 1}. \quad (\text{S88})$$

Combining inequalities (S87) and (S88) gives

$$\kappa\rho^{-2} + m < \sqrt{m} \cdot \frac{\sqrt{m}}{\kappa + 1} = \frac{m}{\kappa + 1} < m,$$

leading to a contradiction. Therefore, it remains to consider scenario (i), i.e., $\sqrt{m}(\kappa\rho^{-2} + m) - \sqrt{2\rho^{-2}}m > 0$ and $\sqrt{2\rho^{-2}}(\kappa + 1) - \sqrt{m} > 0$. From (S83), we can know that $\frac{\partial \tilde{H}_{m,1}(\gamma; c, \rho)}{\partial (\gamma^2)}$ is negative when $\gamma < \gamma_0$ and positive when $\gamma > \gamma_0$. Consequently, $\tilde{H}_{m,1}(\gamma; c, \rho)$ is decreasing for $\gamma < \gamma_0$ and increasing for $\gamma > \gamma_0$.

From the above, we have:

- (a) in the first and second scenarios, $\tilde{H}_{m,1}(\gamma; c, \rho)$ is either increasing, decreasing, or constant in $\gamma \in (\rho^{-1}, \infty)$;
- (b) in the third scenario, $\tilde{H}_{m,1}(\gamma; c, \rho)$ is decreasing for $\gamma \in (\rho^{-1}, \gamma_0)$ and increasing for $\gamma \in (\gamma_0, \infty)$, for some $\gamma_0 > \rho^{-1}$.

This means $\tilde{H}_{m,1}(\gamma; c, \rho)$ is bounded from above by its endpoints, i.e., it is bounded from above by

$$\bar{H}_{m,1}(c, \rho) \equiv \max \left\{ \tilde{H}_{m,1}(\rho^{-1}; c, \rho), \lim_{\gamma \rightarrow \infty} \tilde{H}_{m,1}(\gamma; c, \rho) \right\},$$

where

$$\begin{aligned} \tilde{H}_{m,1}(\rho^{-1}; c, \rho) &= \frac{m + \rho^{-2}}{(\kappa + 1)\rho^{-2} + m} + \frac{2(m + \rho^{-2})}{(\kappa + 1)\rho^{-2} + m} = \frac{3(m + \rho^{-2})}{(\kappa + 1)\rho^{-2} + m}, \\ \lim_{\gamma \rightarrow \infty} \tilde{H}_{m,1}(\gamma; c, \rho) &= \lim_{\gamma \rightarrow \infty} \frac{m + \gamma^2}{(\kappa + 1)\gamma^2 + m} + 2 \lim_{\gamma \rightarrow \infty} \left[\frac{m + \gamma^2}{\gamma^2 + (\kappa\rho^{-2} + m)} \right] \\ &= \frac{1}{\kappa + 1} + 2 = \frac{2\kappa + 3}{\kappa + 1}. \end{aligned}$$

Therefore, we have proved Lemma S2.5. □

Lemma S2.6. For any given positive c, ρ, m and any $\gamma > \rho^{-1}$, define

$$\begin{aligned} D_1(\gamma) &= \frac{1}{2}\gamma^4 + \left(\kappa\rho^{-2} + \frac{1}{2}\rho^{-2} + \frac{1}{2}m \right) \gamma^2 + \frac{1}{2}\rho^{-2}m, \\ D_2(\gamma) &= (\kappa + 2)\gamma^4 + (\kappa m + \kappa\rho^{-2} + 4m)\gamma^2 + m(\kappa\rho^{-2} + 2m), \\ Z(\gamma) &= \frac{D_1(\gamma)}{D_2(\gamma)}. \end{aligned}$$

For all $\gamma > \rho^{-1}$, we have $Z(\gamma) \geq \underline{Z}$, where

$$\underline{Z} \equiv \min \left\{ \frac{1}{2(m\rho^2 + 1)}, \frac{1}{2(\kappa + 2)} \right\} = \frac{1}{2 \cdot \max\{m\rho^2 + 1, \kappa + 2\}} > 0. \quad (\text{S89})$$

Proof of Lemma S2.6. To derive the lower bound on $Z(\gamma)$, we will study $\frac{\partial Z(\gamma)}{\partial \gamma^2}$. The derivatives of $D_1(\gamma)$ and $D_2(\gamma)$ with respect to γ^2 are, respectively:

$$\begin{aligned} \frac{\partial D_1(\gamma)}{\partial \gamma^2} &= \gamma^2 + \kappa\rho^{-2} + \frac{1}{2}(m + \rho^{-2}), \\ \frac{\partial D_2(\gamma)}{\partial \gamma^2} &= 2(\kappa + 2)\gamma^2 + (\kappa m + \kappa\rho^{-2} + 4m). \end{aligned}$$

Using the above, we have

$$\begin{aligned} \frac{\partial Z(\gamma)}{\partial \gamma^2} &= \frac{D_2(\gamma) \frac{\partial D_1(\gamma)}{\partial (\gamma^2)} - D_1(\gamma) \frac{\partial D_2(\gamma)}{\partial (\gamma^2)}}{\{D_2(\gamma)\}^2} \\ &= \frac{[m - \rho^{-2}(\kappa + 1)^2]\gamma^4 + 2m(m - \rho^{-2})\gamma^2 + m[\kappa^2\rho^{-4} + m^2 + m\rho^{-2}(2\kappa - 1)]}{\{D_2(\gamma)\}^2} \\ &= \frac{Z_1\gamma^4 + Z_2\gamma^2 + Z_3}{\{D_2(\gamma)\}^2}, \end{aligned} \quad (\text{S90})$$

where $Z_1 \equiv m - \rho^{-2}(\kappa + 1)^2$, $Z_2 \equiv 2m(m - \rho^{-2})$, and $Z_3 \equiv m[\kappa^2\rho^{-4} + m^2 + m\rho^{-2}(2\kappa - 1)]$. Let $f(\gamma^2) \equiv Z_1\gamma^4 + Z_2\gamma^2 + Z_3$ be a quadratic function of γ^2 . In the following, we consider three different cases depending on the sign of Z_1 .

First, suppose that $Z_1 > 0$. Since $\kappa > 0$, it follows that

$$Z_2 = 2m(m - \rho^{-2}) > 2m[m - \rho^{-2}(\kappa + 1)^2] = 2mZ_1 > 0. \quad (\text{S91})$$

In addition, $Z_1 > 0$ implies that $m\rho^2 > (\kappa + 1)^2$, or equivalently,

$$\kappa < \rho\sqrt{m} - 1. \quad (\text{S92})$$

Note that Z_3 can be written as

$$\begin{aligned} Z_3 &= m[\kappa^2\rho^{-4} + 2m\kappa\rho^{-2} + m^2 - m\rho^{-2}] = m[(\kappa\rho^{-2} + m)^2 - m\rho^{-2}] \\ &= m(\kappa\rho^{-2} + m + \sqrt{m}\rho^{-1})(\kappa\rho^{-2} + m - \sqrt{m}\rho^{-1}) \\ &= m\rho^{-2}(\kappa\rho^{-2} + m + \sqrt{m}\rho^{-1})(\kappa + m\rho^2 - \sqrt{m}\rho) \end{aligned}$$

$$= m\rho^{-2}(\kappa\rho^{-2} + m + \sqrt{m}\rho^{-1})[\kappa + \sqrt{m}\rho(\sqrt{m}\rho - 1)]. \quad (\text{S93})$$

Combining (S92) and (S93), we have

$$Z_3 > m\rho^{-2}(\kappa\rho^{-2} + m + \sqrt{m}\rho^{-1})(\kappa + \sqrt{m}\rho\kappa) > 0. \quad (\text{S94})$$

(S91) and (S94) then imply that $-\frac{Z_2}{Z_1} < 0$ and $\frac{Z_3}{Z_1} > 0$. If the quadratic function $f(\cdot)$ have real roots, then, by Vieta's formulas, the sum of the two roots are negative while the product of roots is positive, implying that both roots of $f(\gamma^2) = 0$ are negative. Otherwise, the quadratic function $f(\cdot)$ takes positive values on the whole real line. In both cases, we have $f(\gamma^2) > 0$ for all $\gamma > 0$.

Second, consider the case where $Z_1 = 0$. In this case, $f(\gamma^2) = Z_2\gamma^2 + Z_3$. By the same argument as in (S91), we have $Z_2 > 0$. Note that $Z_1 = 0$ implies $\kappa = \sqrt{m}\rho - 1$. Using (S93), we have $Z_3 > 0$. Therefore, $f(\gamma^2) > 0$ for all $\gamma > 0$.

Third, consider the case where $Z_1 < 0$. We consider the following cases.

- Suppose that $f(\cdot)$ has real roots and $Z_2 < 0$. Then the sum of the two roots of $f(\cdot)$ equals $-\frac{Z_2}{Z_1} < 0$. Thus, there is at most one positive root of $f(\cdot)$. If $f(\cdot)$ has a positive root γ_0^2 , i.e., $f(\gamma_0^2) = 0$, then $f(\gamma^2)$ must be positive when $0 < \gamma < \gamma_0$ and negative when $\gamma > \gamma_0$. Otherwise, $f(\gamma^2) < 0$ for all $\gamma > 0$.
- Suppose that $f(\cdot)$ has real roots and $Z_2 \geq 0$. By definition, this means $\sqrt{m} \geq \rho^{-1}$, or equivalently, $\sqrt{m}\rho - 1 \geq 0$. From (S93), because $\kappa > 0$, this implies that $Z_3 > 0$. Thus, the product of the two roots of $f(\cdot)$ equals $\frac{Z_3}{Z_1} < 0$. Consequently, $f(\cdot)$ has one positive root and one negative root. Let γ_0^2 be the positive root of $f(\cdot)$, i.e., $f(\gamma_0^2) = 0$. Then $f(\gamma^2)$ must be positive when $0 < \gamma < \gamma_0$ and negative when $\gamma > \gamma_0$.
- Suppose that $f(\cdot)$ does not have real roots. Then the quadratic function $f(\cdot)$ takes negative values on the whole real line.

From the above, one of the following must hold for $f(\cdot)$:

- (i) $f(\gamma^2) > 0$ for all $\gamma > 0$;
- (ii) $f(\gamma^2) < 0$ for all $\gamma > 0$;
- (iii) for some $\gamma_0 > 0$, $f(\gamma^2) > 0$ for $0 < \gamma < \gamma_0$, $f(\gamma_0^2) = 0$, and $f(\gamma^2) < 0$ for $\gamma > \gamma_0$.

From (S90), we can know that

- (i) $Z(\gamma)$ is either increasing or decreasing in $\gamma \in (\rho^{-1}, \infty)$;

- (ii) for some $\gamma_0 > \rho^{-1}$, $Z(\gamma)$ is increasing for $\gamma \in (\rho^{-1}, \gamma_0)$ and decreasing for $\gamma \in (\gamma_0, \infty)$

Thus, the infimum of $Z(\gamma)$ over $\gamma \in (\rho^{-1}, \infty)$ must be obtained at the endpoints, which immediately implies that, for all $\gamma > \rho^{-1}$,

$$Z(\gamma) \geq \min \left\{ Z(\rho^{-1}), \lim_{\gamma \rightarrow \infty} Z(\gamma) \right\}, \quad (\text{S95})$$

where

$$Z(\rho^{-1}) = \frac{\rho^{-4}(\kappa + \rho^2 m + 1)}{2\rho^{-4}(m\rho^2 + 1)(\kappa + m\rho^2 + 1)} = \frac{1}{2(m\rho^2 + 1)},$$

$$\lim_{\gamma \rightarrow \infty} Z(\gamma) = \frac{1}{2(\kappa + 2)}.$$

By definition, this then implies that $Z(\gamma) \geq \underline{Z}$ for all $\gamma > \rho^{-1}$. Moreover, we have $\underline{Z} > 0$ because $\kappa, m, \rho > 0$. Therefore, we derive Lemma S2.6. \square

Lemma S2.7. For any given values of positive $m \geq 4$, $c > \sqrt{\frac{2(m-1)}{m(m-2)}}$ and ρ , define $\tilde{H}_{m,2}(\gamma; c, \rho)$ as in (S78) for $\gamma > \rho^{-1}$. For any $\gamma > \rho^{-1}$, we have $\tilde{H}_{m,2}(\gamma; c, \rho) \leq \bar{H}_{m,2}(c, \rho)$, where

$$\bar{H}_{m,2}(c, \rho) \equiv \frac{1 - \tau}{1 - \tau + \min \left\{ \left[(1 - 2\tau)\kappa\underline{Z} - \frac{1}{2} \right], 0 \right\}}, \quad (\text{S96})$$

and \underline{Z} is defined as in (S89).

Proof of Lemma S2.7. Recall that $\tilde{H}_{m,2}(\gamma; c, \rho) \equiv \frac{1-\tau}{1-\tau+\min\{C,0\}}$ with

$$C \equiv \left(1 - 2\tau + \frac{2}{\kappa\gamma^2} \right) \frac{\kappa^2\gamma^2\rho^{-2} - \bar{\theta}^2}{(\kappa\gamma^2 + \kappa\rho^{-2} + 2\bar{\theta})\bar{\theta}} + \frac{1}{\kappa\gamma^2} - \tau \quad \text{and} \quad \bar{\theta} = m + \gamma^2.$$

Define $D_1(\gamma)$ and $D_2(\gamma)$ as in (S2.6). We have

$$\begin{aligned} (\kappa\gamma^2 + \kappa\rho^{-2} + 2\bar{\theta})\bar{\theta} &= [(\kappa + 2)\gamma^2 + 2m + \kappa\rho^{-2}](\gamma^2 + m) \\ &= D_2(\gamma), \end{aligned}$$

and

$$(\kappa^2\gamma^2\rho^{-2} - \bar{\theta}^2) + \frac{1}{2}(\kappa\gamma^2 + \kappa\rho^{-2} + 2\bar{\theta})\bar{\theta} = \kappa^2\gamma^2\rho^{-2} + \frac{1}{2}\kappa\gamma^2\bar{\theta} + \frac{1}{2}\kappa\rho^{-2}\bar{\theta} \equiv D_1(\gamma) \cdot \kappa > 0. \quad (\text{S97})$$

By the same logic as the proof of Lemma A.3, because $c \geq \sqrt{\frac{2(m-1)}{m(m-2)}}$, we have $\tau < 1/2$. Hence, we have

$$\begin{aligned} C &= \left(1 - 2\tau + \frac{2}{\kappa\gamma^2}\right) \frac{D_1(\gamma)\kappa - \frac{1}{2}D_2(\gamma)}{D_2(\gamma)} + \frac{1}{\kappa\gamma^2} - \tau = \left(1 - 2\tau + \frac{2}{\kappa\gamma^2}\right) \frac{D_1(\gamma)\kappa}{D_2(\gamma)} - \frac{1}{2} \\ &\geq (1 - 2\tau) \frac{D_1(\gamma)\kappa}{D_2(\gamma)} - \frac{1}{2} = (1 - 2\tau)\kappa Z(\gamma) - \frac{1}{2} \geq (1 - 2\tau)\kappa \underline{Z} - \frac{1}{2}, \end{aligned} \quad (\text{S98})$$

where the last inequality follows from Lemma S2.6 and that $\tau < 1/2$. This then implies that

$$1 - \tau + \min\{C, 0\} \geq 1 - \tau + \min\left\{\left[(1 - 2\tau)\kappa \underline{Z} - \frac{1}{2}\right], 0\right\} > 0,$$

and consequently

$$\tilde{H}_{m,2}(\gamma; c, \rho) = \frac{1 - \tau}{1 - \tau + \min\{C, 0\}} \leq \frac{1 - \tau}{1 - \tau + \min\left\{\left[(1 - 2\tau)\kappa \underline{Z} - \frac{1}{2}\right], 0\right\}} = \bar{H}_{m,2}(c, \rho). \quad (\text{S99})$$

Therefore, Lemma S2.7 holds. \square

Lemma S2.8. For any given values of positive c , ρ , and $m \geq 2$, define $\bar{H}_{m,1}(c, \rho)$ as in (S86). We have that $\bar{H}_{m,1}(c, \rho)$ is decreasing in $c > 0$.

Proof of Lemma S2.8. By definition,

$$\bar{H}_{m,1}(c, \rho) = \max\left\{\frac{3(m + \rho^{-2})}{(\kappa + 1)\rho^{-2} + m}, \frac{2\kappa + 3}{\kappa + 1}\right\},$$

where $\kappa = \frac{mc^2}{m-1}$ is increasing in $c > 0$. Note that both $\frac{3(m + \rho^{-2})}{(\kappa + 1)\rho^{-2} + m}$ and $\frac{2\kappa + 3}{\kappa + 1} = 2 + \frac{1}{\kappa + 1}$ are decreasing in κ . These then imply that $\bar{H}_{m,1}(c, \rho)$ is decreasing in κ . Consequently, $\bar{H}_{m,1}(c, \rho)$ is decreasing in $c > 0$. Therefore, Lemma S2.8 holds. \square

Lemma S2.9. For any given values of positive $m \geq 4$, $c > \sqrt{\frac{2(m-1)}{m(m-2)}}$ and ρ , define $\bar{H}_{m,2}(c, \rho)$ as in (S96). We have that $\bar{H}_{m,2}(c, \rho)$ is nonincreasing in c .

Proof of Lemma S2.9. Because κ is increasing in $c > 0$, as shown in the proof of Lemma S2.8, it suffices to prove that $\bar{H}_{m,2}(c, \rho)$ is nonincreasing in κ . By definition,

$$\bar{H}_{m,2}(c, \rho) = \frac{1 - \tau}{1 - \tau + \min\left\{(1 - 2\tau)\kappa \underline{Z} - \frac{1}{2}, 0\right\}},$$

where

$$\kappa = \frac{mc^2}{m-1}, \quad \tau = \frac{\kappa+1}{m\kappa}, \quad \underline{Z} = \min \left\{ \frac{1}{2(m\rho^2+1)}, \frac{1}{2(\kappa+2)} \right\}.$$

Note that

$$(1-2\tau)\kappa\underline{Z} - \frac{1}{2} = (1-2\tau) \min \left\{ \frac{\kappa}{2(m\rho^2+1)}, \frac{\kappa}{2(\kappa+2)} \right\} - \frac{1}{2}.$$

Obviously, $\frac{\kappa}{2(m\rho^2+1)}$ and $\frac{\kappa}{2(\kappa+2)}$ are increasing in κ , which implies that $\kappa\underline{Z}$ is increasing in κ . In addition, $\tau = \frac{\kappa+1}{m\kappa} = \frac{1}{m} + \frac{1}{m\kappa}$ is decreasing in κ . Thus, $(1-2\tau)\kappa\underline{Z} - \frac{1}{2}$ is increasing in κ . Below we consider two cases depending on whether there exists $\kappa_0 > 0$ such that $(1-2\tau)\kappa\underline{Z} - \frac{1}{2} = 0$ when evaluated at $\kappa = \kappa_0$.

First, assume that such $\kappa_0 > 0$ exists. Since $(1-2\tau)\kappa\underline{Z} - \frac{1}{2}$ is increasing in κ , we can know that $(1-2\tau)\kappa\underline{Z} - \frac{1}{2} < 0$ for $\kappa < \kappa_0$ and $(1-2\tau)\kappa\underline{Z} - \frac{1}{2} > 0$ for $\kappa > \kappa_0$. Thus, for $\kappa \geq \kappa_0$,

$$\bar{H}_{m,2}(c, \rho) = \frac{1-\tau}{1-\tau} = 1, \quad (\text{S100})$$

which is a constant over all $\kappa \geq \kappa_0$. For $\kappa < \kappa_0$,

$$\bar{H}_{m,2}(c, \rho) = \frac{1-\tau}{1-\tau + \left[(1-2\tau)\kappa\underline{Z} - \frac{1}{2} \right]} = \frac{2(1-\tau)}{(1-2\tau)(1+2\kappa\underline{Z})} = \frac{1}{1+2\kappa\underline{Z}} \left(1 + \frac{1}{1-2\tau} \right). \quad (\text{S101})$$

From the discussion before, $\kappa\underline{Z}$ is increasing in κ , and τ is decreasing in κ . These imply that (S101) must be decreasing in $\kappa \in (0, \kappa_0)$. Therefore, $\bar{H}_{m,2}(c, \rho)$ is nonincreasing in κ .

Second, assume that such κ does not exist. Then $(1-2\tau)\kappa\underline{Z} - \frac{1}{2}$ is either positive for any $\kappa > 0$, or negative for any $\kappa > 0$. Below we consider the two cases, separately.

- Suppose that $(1-2\tau)\kappa\underline{Z} - \frac{1}{2} > 0$ for any $\kappa > 0$. Then $\bar{H}_{m,2}(c, \rho)$ has the form in (S100). This implies that $\bar{H}_{m,2}(c, \rho)$ is constant over all $\kappa > 0$.
- Suppose that $(1-2\tau)\kappa\underline{Z} - \frac{1}{2} < 0$ for any $\kappa > 0$. Then $\bar{H}_{m,2}(c, \rho)$ has the form in (S101). For the same reason as in the first case, $\bar{H}_{m,2}(c, \rho)$ is decreasing in $\kappa > 0$.

From the above, Lemma S2.9 holds. \square

Lemma S2.10. For given values of positive c , ρ , and $m \geq 2$, define $\tilde{H}_{m,3}(c, \rho)$ as in (S78). We have that $\tilde{H}_{m,3}(c, \rho)$ is decreasing in $c > 0$

Proof of Lemma S2.10. Because κ is increasing in $c > 0$, as shown in the proof of Lemma

S2.8, it suffices to prove that $\tilde{H}_{m,3}(c, \rho)$ is decreasing in κ . By definition,

$$\frac{\partial}{\partial \kappa} \tilde{H}_{m,3}(c, \rho) \equiv -(\underline{\theta} - m) \cdot \frac{\partial}{\partial \kappa} \left(\frac{1}{\tau \underline{\theta} - 1} \right) = (\underline{\theta} - m) \cdot \frac{\underline{\theta}}{(\tau \underline{\theta} - 1)^2} \frac{\partial \tau}{\partial \kappa},$$

where $\underline{\theta} = m + \rho^{-2}$ and $\tau = \frac{\kappa+1}{m\kappa}$. Because $\underline{\theta} - m = \rho^{-2} > 0$, $\underline{\theta} > 0$, and $\tau = \frac{1}{m} + \frac{1}{m\kappa}$ is decreasing in κ , we can know that $\frac{\partial}{\partial \kappa} \tilde{H}_{m,3}(c, \rho) < 0$. Consequently, $\tilde{H}_{m,3}(c, \rho)$ is decreasing in κ . Therefore, Lemma S2.10 holds. \square

Proof of Theorem 4.4. Define $\tilde{H}_m(\gamma; c, \rho)$ as in Theorem S1, $\tilde{H}_{m,1}(\gamma; c, \rho)$, $\tilde{H}_{m,2}(\gamma; c, \rho)$ and $\tilde{H}_{m,3}(c, \rho)$ as in (S78), $\bar{H}_{m,1}(c, \rho)$ as in Lemma S2.5, and $\bar{H}_{m,2}(c, \rho)$ as in Lemma S2.7. By definition and from Lemmas S2.5 and S2.7, for any $c > \sqrt{\frac{2(m-1)}{m(m-2)}}$,

$$\begin{aligned} \tilde{H}_m(\gamma; c, \rho) &= \tilde{H}_{m,1}(\gamma; c, \rho) + \tilde{H}_{m,2}(\gamma; c, \rho) + \tilde{H}_{m,3}(c, \rho) - 1 \\ &\leq \bar{H}_{m,1}(c, \rho) + \bar{H}_{m,2}(c, \rho) + \tilde{H}_{m,3}(c, \rho) - 1. \end{aligned} \quad (\text{S102})$$

In addition, by definition,

$$\begin{aligned} &\bar{H}_{m,1}(c, \rho) + \bar{H}_{m,2}(c, \rho) + \tilde{H}_{m,3}(c, \rho) - 1 \\ &= \max \left\{ \frac{3(m + \rho^{-2})}{(\kappa + 1)\rho^{-2} + m}, \frac{2\kappa + 3}{\kappa + 1} \right\} + \frac{1 - \tau}{1 - \tau + \min\{(1 - 2\tau)\kappa \underline{Z} - \frac{1}{2}, 0\}} - \frac{\underline{\theta} - m}{\tau \underline{\theta} - 1} - 1 \\ &= \max \left\{ \frac{3(m\rho^2 + 1)}{m\rho^2 + \kappa + 1}, \frac{2\kappa + 3}{\kappa + 1} \right\} + \frac{1 - \tau}{1 - \tau + \min\{(1 - 2\tau)\kappa \underline{Z} - \frac{1}{2}, 0\}} - \frac{\rho^{-2}}{\tau m + \tau \rho^{-2} - 1} - 1 \\ &= \max \left\{ \frac{3(m\rho^2 + 1)}{m\rho^2 + \kappa + 1}, \frac{2\kappa + 3}{\kappa + 1} \right\} + \frac{1 - \tau}{1 - \tau + \min\{(1 - 2\tau)\kappa \underline{Z} - \frac{1}{2}, 0\}} - \frac{m\kappa}{m\rho^2 + \kappa + 1} - 1 \\ &= \bar{H}_m(c, \rho), \end{aligned} \quad (\text{S103})$$

where κ , τ , $\underline{\theta}$ is defined as in (S79), \underline{Z} is defined as in (S89), the first and last equalities follow by definition, and the second and third equalities follow by some algebra.

First, from Lemmas S2.8–S2.10, $\bar{H}_m(c, \rho) = \bar{H}_{m,1}(c, \rho) + \bar{H}_{m,2}(c, \rho) + \tilde{H}_{m,3}(c, \rho) - 1$ is decreasing in c . Moreover, as $c \rightarrow \infty$, we have $\kappa = mc^2/(m-1) \rightarrow \infty$, $\tau = (\kappa + 1)/(m\kappa) \rightarrow 1/m$,

$$\kappa \underline{Z} = \frac{\kappa}{2 \cdot \max\{m\rho^2 + 1, \kappa + 2\}} \rightarrow \frac{1}{2},$$

and consequently

$$\begin{aligned}
& \bar{H}_m(c, \rho) \\
&= \max \left\{ \frac{3(m\rho^2 + 1)}{m\rho^2 + \kappa + 1}, \frac{2\kappa + 3}{\kappa + 1} \right\} + \frac{1 - \tau}{1 - \tau + \min\{(1 - 2\tau)\kappa\underline{Z} - \frac{1}{2}, 0\}} - \frac{m\kappa}{m\rho^2 + \kappa + 1} - 1 \\
&\rightarrow \max\{0, 2\} + \frac{1 - m^{-1}}{1 - m^{-1} + \min\{(1 - 2m^{-1})\frac{1}{2} - \frac{1}{2}, 0\}} - m - 1 \\
&= 1 - m + \frac{1 - m^{-1}}{1 - m^{-1} - m^{-1}} = 1 - m + \frac{m - 1}{m - 2} = \frac{(1 - m)(m - 3)}{m - 2} < 0, \tag{S104}
\end{aligned}$$

where the last inequality holds because $m \geq 4$.

Second, by definition and from (S104), we know that $\underline{c}_{m,\rho} = \inf\{c > \sqrt{\frac{3(m-1)}{m(m-3)}} : \bar{H}_m(c, \rho) \leq 0\}$ must be finite. Because $\bar{H}_m(c, \rho)$ is decreasing in c , we must have $\bar{H}_m(c, \rho) < 0$ for any $c > \underline{c}_{m,\rho}$. From (S102) and (S103), we then have, for any given $m \geq 4, \rho > 0, c > \underline{c}_{m,\rho} \geq \sqrt{\frac{3(m-1)}{m(m-3)}} > \sqrt{\frac{2(m-1)}{m(m-2)}}$,

$$\tilde{H}_m(\gamma; c, \rho) \leq \bar{H}_m(c, \rho) < 0 \quad \text{for all } \gamma > \rho^{-1}.$$

From Theorem S1, this implies that, for any given $m \geq 4, \rho > 0, c > \underline{c}_{m,\rho}$, the maximum rejection probability $p_m(c; 1, \rho)$ under Assumption 3.2 with $k = 1$ and the given ρ, m , and c has the following equivalent form:

$$p_m(c; 1, \rho) = \mathbb{P} \left[|t_{m-1}| \sqrt{\rho^2 + \frac{1}{m}} > c \right].$$

Third, we consider the case when $c = \underline{c}_{m,\rho}$. From Lemma A.2, there exist $\{\sigma_1, \dots, \sigma_{m+1}\} \in \mathcal{S}_m(1, \rho)$ such that $\mathbb{P}_0[|T_m| > \underline{c}_{m,\rho}] = p_m(\underline{c}_{m,\rho}; 1, \rho)$. By the right continuity of distribution functions, we can know

$$\begin{aligned}
p_m(\underline{c}_{m,\rho}; 1, \rho) &= \mathbb{P}_0[|T_m| > \underline{c}_{m,\rho}] \\
&= \lim_{\epsilon \rightarrow 0+} \mathbb{P}_0[|T_m| > \underline{c}_{m,\rho} + \epsilon] \\
&\leq \lim_{\epsilon \rightarrow 0+} \mathbb{P} \left[|t_{m-1}| \sqrt{\rho^2 + \frac{1}{m}} > \underline{c}_{m,\rho} + \epsilon \right] \\
&= \mathbb{P} \left[|t_{m-1}| \sqrt{\rho^2 + \frac{1}{m}} > \underline{c}_{m,\rho} \right], \tag{S105}
\end{aligned}$$

where the last inequality follows from the second part. Moreover, the rejection probability can reach the upper bound on the right-hand side of (S105) when $\sigma_{m+1} = \rho$ and $\sigma_1 = \dots = \sigma_m = 1$. Thus, (S105) must hold with equality.

From the above, Theorem 4.4 holds. \square

S2.7 Proof of Theorem 4.5

Proof of Theorem 4.5. Because $c > 0$, we obviously have $\mathbb{P}[|T_m| > c] \geq \mathbb{P}[T_m > c] = 1 - \mathbb{P}[T_m \leq c]$. Note that

$$\begin{aligned} T_m \leq c &\iff (\psi_{m+1} - \delta) + \delta - \bar{\psi} \leq cS_m \\ &\iff cS_m + \bar{\psi} - (\psi_{m+1} - \delta) \geq \delta \\ &\implies \{cS_m + \bar{\psi} - (\psi_{m+1} - \delta)\}^2 \geq \delta^2, \end{aligned}$$

where the last step holds because $\delta > 0$. By the Markov inequality, we then have

$$\mathbb{P}[T_m \leq c] \leq \mathbb{P}[\{cS_m + \bar{\psi} - (\psi_{m+1} - \delta)\}^2 \geq \delta^2] \leq \delta^{-2} \mathbb{E}[\{cS_m + \bar{\psi} - (\psi_{m+1} - \delta)\}^2].$$

Under Assumption 4.1, we have

$$\begin{aligned} \mathbb{E}[\{cS_m + \bar{\psi} - (\psi_{m+1} - \delta)\}^2] &= \mathbb{E}[(cS_m + \bar{\psi})^2] + \mathbb{E}[(\psi_{m+1} - \delta)^2] \\ &\leq 2\mathbb{E}[c^2 S_m^2] + 2\mathbb{E}[\bar{\psi}^2] + \mathbb{E}[(\psi_{m+1} - \delta)^2], \end{aligned}$$

where the first equality holds because $\psi_{m+1} - \delta \sim \mathcal{N}(0, \sigma_{m+1}^2)$ and is independent of $\{\psi_j\}_{j=1}^m$, and the second equality follows from the Cauchy-Schwarz inequality. By some algebra, $\mathbb{E}[(\psi_{m+1} - \delta)^2] = \text{Var}(\psi_{m+1}) = \sigma_{m+1}^2$, $\mathbb{E}[\bar{\psi}^2] = \text{Var}(\bar{\psi}) = \frac{1}{m^2} \sum_{j=1}^m \sigma_j^2$,

$$\mathbb{E}[S_m^2] = \frac{1}{m-1} \sum_{j=1}^m \mathbb{E}[\psi_j^2] - \frac{m}{m-1} \mathbb{E}[\bar{\psi}^2] = \frac{1}{m-1} \sum_{j=1}^m \sigma_j^2 - \frac{m}{m-1} \frac{1}{m^2} \sum_{j=1}^m \sigma_j^2 = \frac{1}{m} \sum_{j=1}^m \sigma_j^2.$$

These then imply that

$$\begin{aligned} \mathbb{E}[\{cS_m + \bar{\psi} - (\psi_{m+1} - \delta)\}^2] &\leq 2\mathbb{E}[c^2 S_m^2] + 2\mathbb{E}[\bar{\psi}^2] + \mathbb{E}[(\psi_{m+1} - \delta)^2] \\ &= 2c^2 \frac{1}{m} \sum_{j=1}^m \sigma_j^2 + \frac{2}{m^2} \sum_{j=1}^m \sigma_j^2 + \sigma_{m+1}^2 \\ &= \sigma_{m+1}^2 + \frac{2(c^2 + m^{-1})}{m} \sum_{j=1}^m \sigma_j^2. \end{aligned}$$

From the above, we have

$$\mathbb{P}[T_m \leq c] \leq \delta^{-2} \mathbb{E}[\{cS_m + \bar{\psi} - (\psi_{m+1} - \delta)\}^2] \leq \frac{1}{\delta^2} \left[\sigma_{m+1}^2 + \frac{2(c^2 + m^{-1})}{m} \sum_{j=1}^m \sigma_j^2 \right],$$

and consequently

$$\mathbb{P}[|T_m| > c] \geq \mathbb{P}[T_m > c] = 1 - \mathbb{P}[T_m \leq c] \geq 1 - \frac{1}{\delta^2} \left[\sigma_{m+1}^2 + \frac{2(c^2 + m^{-1})}{m} \sum_{j=1}^m \sigma_j^2 \right].$$

Therefore, Theorem 4.5 holds. \square

Comment on the number of one-dimensional optimizations needed. We first count the number of one-dimensional optimizations needed at a given k . From Theorem 4.3, we need to consider the one-dimensional optimization in (16) for $0 \leq m_0 \leq k-1$ and $0 \leq m_1 \leq m - m_0 - 1$; note that when $m_1 = m - m_0$ no optimization is needed. Thus, the number of one-dimensional optimization needed is

$$\sum_{m_0=0}^{k-1} (m - m_0) = \sum_{j=m-k+1}^m j = \frac{1}{2}k(2m + 1 - k).$$

We then count the total number of one-dimensional optimizations needed to compute $p_m(c; k, \rho)$ for all $1 \leq k \leq m$. For each combination of (m_1, m_0, m_c) satisfying $m_1 \geq 0$, $m_0 \geq 0$, $m_c \geq 1$, and $m_1 + m_0 + m_c = m$, we solve the one-dimensional optimization in (16) at most twice. Intuitively, m_1 , m_0 , and m_c correspond to the numbers of $\gamma_{j=1}^m$ that are equal to ρ^{-1} , 0, and a common value γ , respectively. The optimization is solved at most twice because γ may vary from 0 or ρ^{-1} to infinity. By some algebra, the total number of one-dimensional optimizations needed is at most

$$2 \cdot \binom{m+1}{2} = m(m+1).$$

\square

S2.8 Proof of Theorem 5.1

Proof of Theorem 5.1. By the definition of $p_m(c; k, \rho)$ in (11), we know that it is non-decreasing in ρ . Thus, the set $\mathcal{I}_{m, \alpha, k}$ in (19) must be an one-sided interval. To prove Theorem 5.1, it suffices to prove the simultaneous validity of the confidence intervals

$\mathcal{I}_{m,\alpha,k}$ s as stated in (ii). Let $G_0(c) = \mathbb{P}_0[|T_m| > c]$ denote the true tail probability of $|T_m|$ at the true values of $\{\sigma_j\}_{j=1}^{m+1}$.

Suppose that $\rho_k^* \notin \mathcal{I}_{m,\alpha,k}$ for some $k \in \{1, 2, \dots, m\}$. By definition, we then have $p_m(|T_m|; k, \rho_k^*) \leq \alpha$ and $\rho_k^* < \infty$. By the definition of ρ_k^* in (18), we must have $\sigma_{m+1} \leq \rho_k^* \sigma_{(k)}$ for the true standard deviation of the treated cluster and that of the control cluster at rank k . By the definition in (11), this then implies that $p_m(c; k, \rho_k^*) \geq G_0(c)$ for any $c \geq 0$. Consequently, we must have $\alpha \geq p_m(|T_m|; k, \rho_k^*) \geq G_0(|T_m|)$.

From the discussion before, we then have

$$\mathbb{P}[\rho_k^* \notin \mathcal{I}_{m,\alpha,k} \text{ for some } 1 \leq k \leq m] \leq \mathbb{P}[G_0(|T_m|) \leq \alpha] \leq \alpha,$$

where the last inequality holds because $G_0(\cdot)$ is the tail probability function of $|T_m|$ (see, e.g., Lemma A4 in Wu and Li (2025) for a proof). This then implies that

$$\mathbb{P}[\rho_k^* \in \mathcal{I}_{m,\alpha,k} \text{ for all } 1 \leq k \leq m] = 1 - \mathbb{P}[\rho_k^* \notin \mathcal{I}_{m,\alpha,k} \text{ for some } 1 \leq k \leq m] \geq 1 - \alpha,$$

Therefore, we derive Theorem 5.1. □

S2.9 Proof of Theorem A.1

Proof of Theorem A.1. From Condition A.1, as $m \rightarrow \infty$,

$$\text{Var}(\bar{\psi}) = \frac{1}{m^2} \sum_{j=1}^m \sigma_j^2 = \frac{1}{m} \cdot \frac{1}{m} \sum_{j=1}^m \sigma_j^2 \rightarrow 0,$$

and

$$\text{Var}\left(\frac{1}{m} \sum_{j=1}^m \psi_j^2\right) = \frac{1}{m^2} \sum_{j=1}^m \text{Var}(\psi_j^2) = \frac{1}{m^2} \sum_{j=1}^m (2\sigma_j^4) = \frac{2}{m^2} \sum_{j=1}^m \sigma_j^4 \rightarrow 0,$$

where the second last equality uses the fact that the variance of a chi-squared random variable with degree of freedom 1 is 2. By Chebyshev's inequality, these then imply that

$$\bar{\psi} = o_{\mathbb{P}}(1), \quad \frac{1}{m} \sum_{j=1}^m \psi_j^2 = \mathbb{E}\left[\frac{1}{m} \sum_{j=1}^m \psi_j^2\right] + o_{\mathbb{P}}(1) = \frac{1}{m} \sum_{j=1}^m \sigma_j^2 + o_{\mathbb{P}}(1).$$

and consequently

$$S_m^2 = \frac{m}{m-1} \frac{1}{m} \sum_{j=1}^m \psi_j^2 - \frac{m}{m-1} \bar{\psi}^2 = \frac{1}{m} \sum_{j=1}^m \sigma_j^2 + o_{\mathbb{P}}(1).$$

These further imply that

$$T_m = \frac{\psi_{m+1} - o_{\mathbb{P}}(1)}{\sqrt{m^{-1} \sum_{j=1}^m \sigma_j^2 + o_{\mathbb{P}}(1)}} = \frac{\psi_{m+1}}{\sqrt{m^{-1} \sum_{j=1}^m \sigma_j^2}} + o_{\mathbb{P}}(1),$$

where the last equality holds under Condition [A.1](#). Let $\varepsilon \equiv \frac{\psi_{m+1} - \delta}{\sigma_{m+1}} \sim \mathcal{N}(0, 1)$. Note that

$$\frac{\psi_{m+1}}{\sqrt{m^{-1} \sum_{j=1}^m \sigma_j^2}} = \frac{\sigma_{m+1}\varepsilon + \delta}{\sqrt{m^{-1} \sum_{j=1}^m \sigma_j^2}} = \frac{\sigma_{m+1}\varepsilon}{\sqrt{m^{-1} \sum_{j=1}^m \sigma_j^2}} + \frac{\delta}{\sqrt{m^{-1} \sum_{j=1}^m \sigma_j^2}}.$$

Using [Wang and Li \(Lemma A27 2022\)](#), we can then derive that, as $m \rightarrow \infty$,

$$\sup_{c \in \mathbb{R}} \left| \mathbb{P}[T_m \leq c] - \mathbb{P}\left[\frac{\sigma_{m+1}\varepsilon + \delta}{\sqrt{m^{-1} \sum_{j=1}^m \sigma_j^2}} \leq c\right] \right| \rightarrow 0.$$

Therefore, Theorem [A.1](#) holds. □

S3 Supplemental results for Section 6

S3.1 Supplemental results for Section 6.1

This section reports additional results from simulations on the first simulation design with $k = 1$ and $k = 2$ at various significance levels.

S3.1.1 Results for $k = 1$

This subsection reports the simulation results for $k = 1$ that are not contained in the main paper. Figures [S1](#) and [S2](#) report the results at the 1% and 10% levels respectively for **DGP 1**. Figures [S3](#) and [S4](#) report the results at the 1% and 10% levels respectively for **DGP 2**.

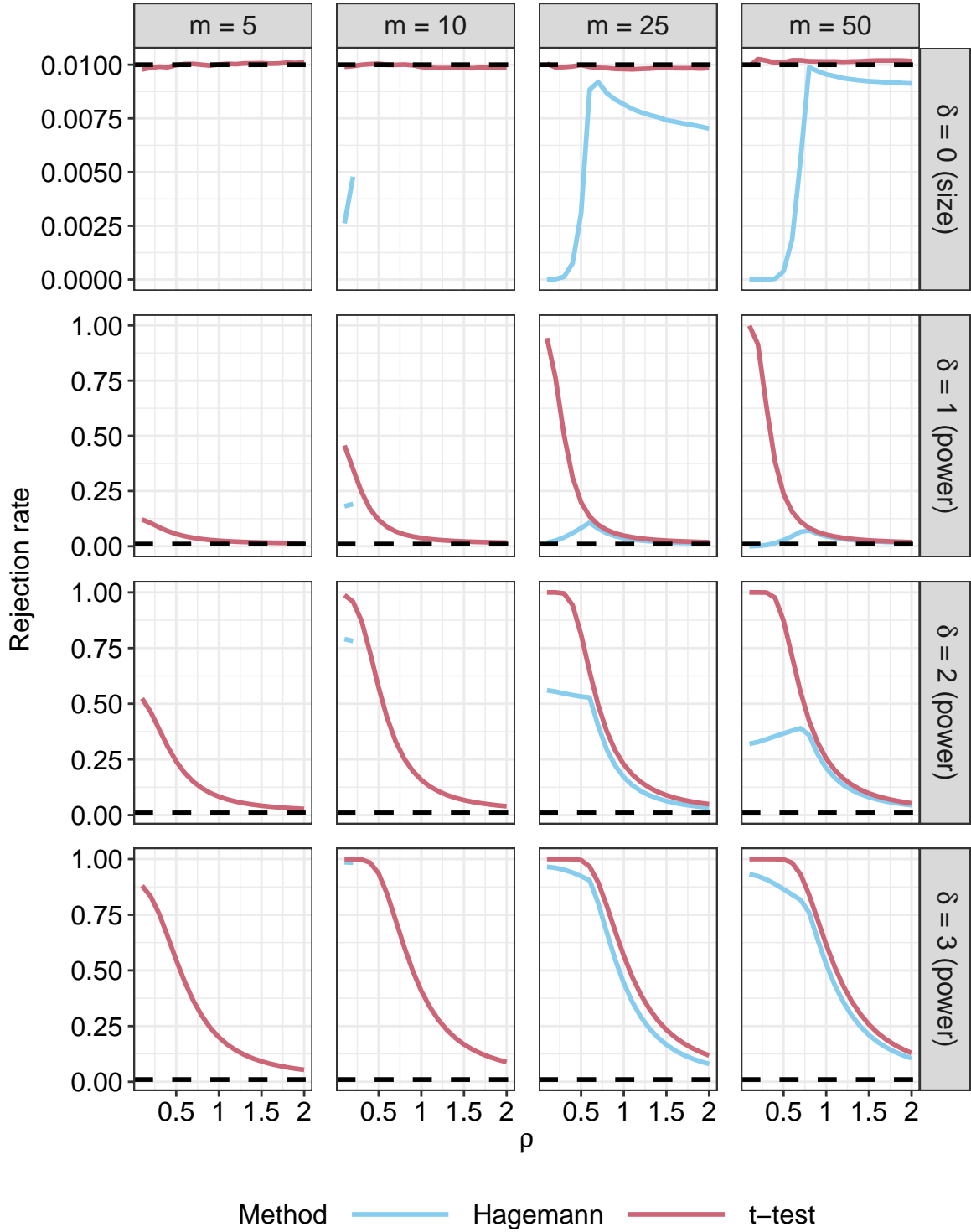


Figure S1: Probability of rejection against heterogeneity parameter ρ for various cluster size m and alternatives δ at $\alpha = 0.01$ for **DGP 1** of simulation design 1 with $k = 1$.

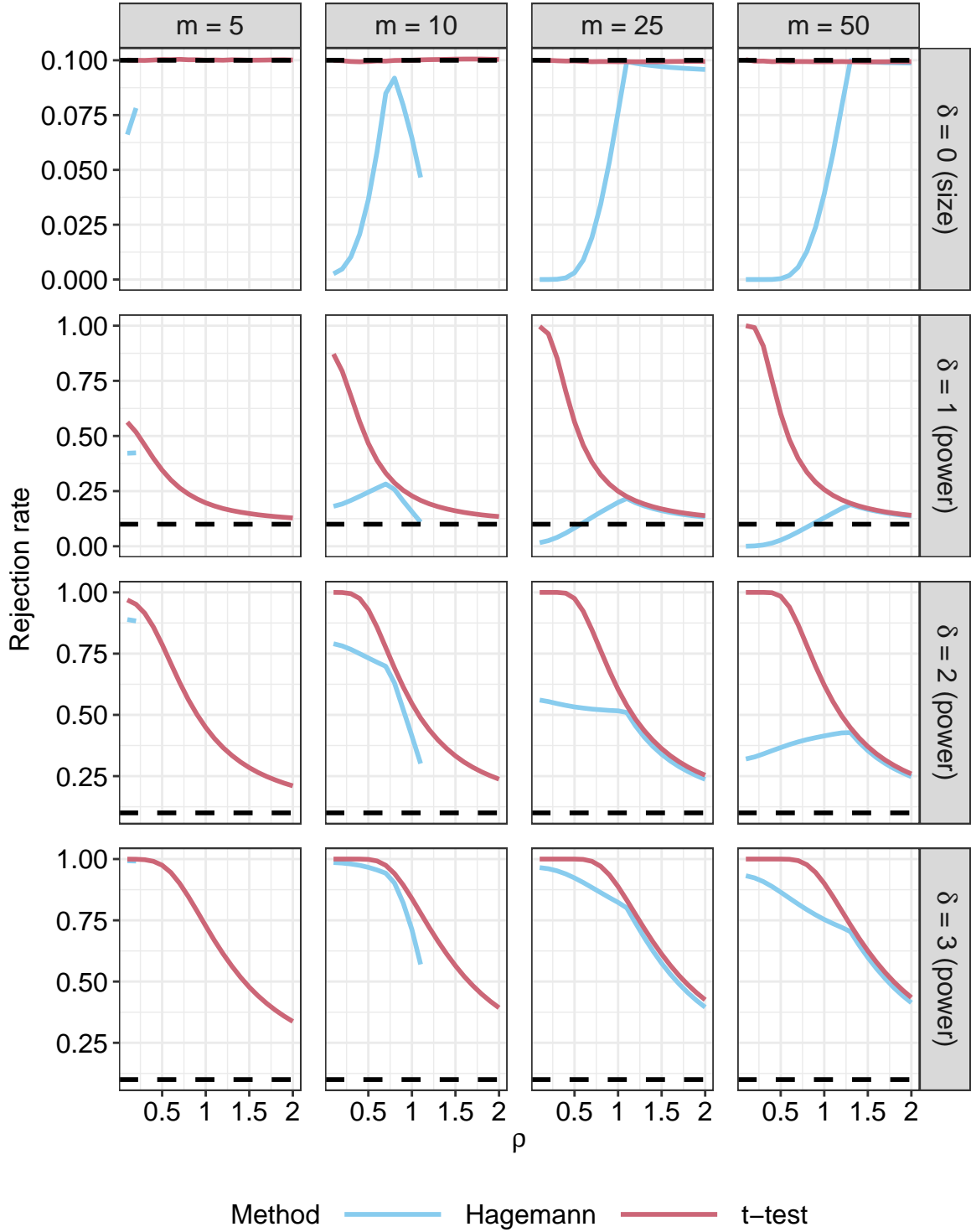


Figure S2: Probability of rejection against heterogeneity parameter ρ for various cluster size m and alternatives δ at $\alpha = 0.1$ for **DGP 1** of simulation design 1 with $k = 1$.

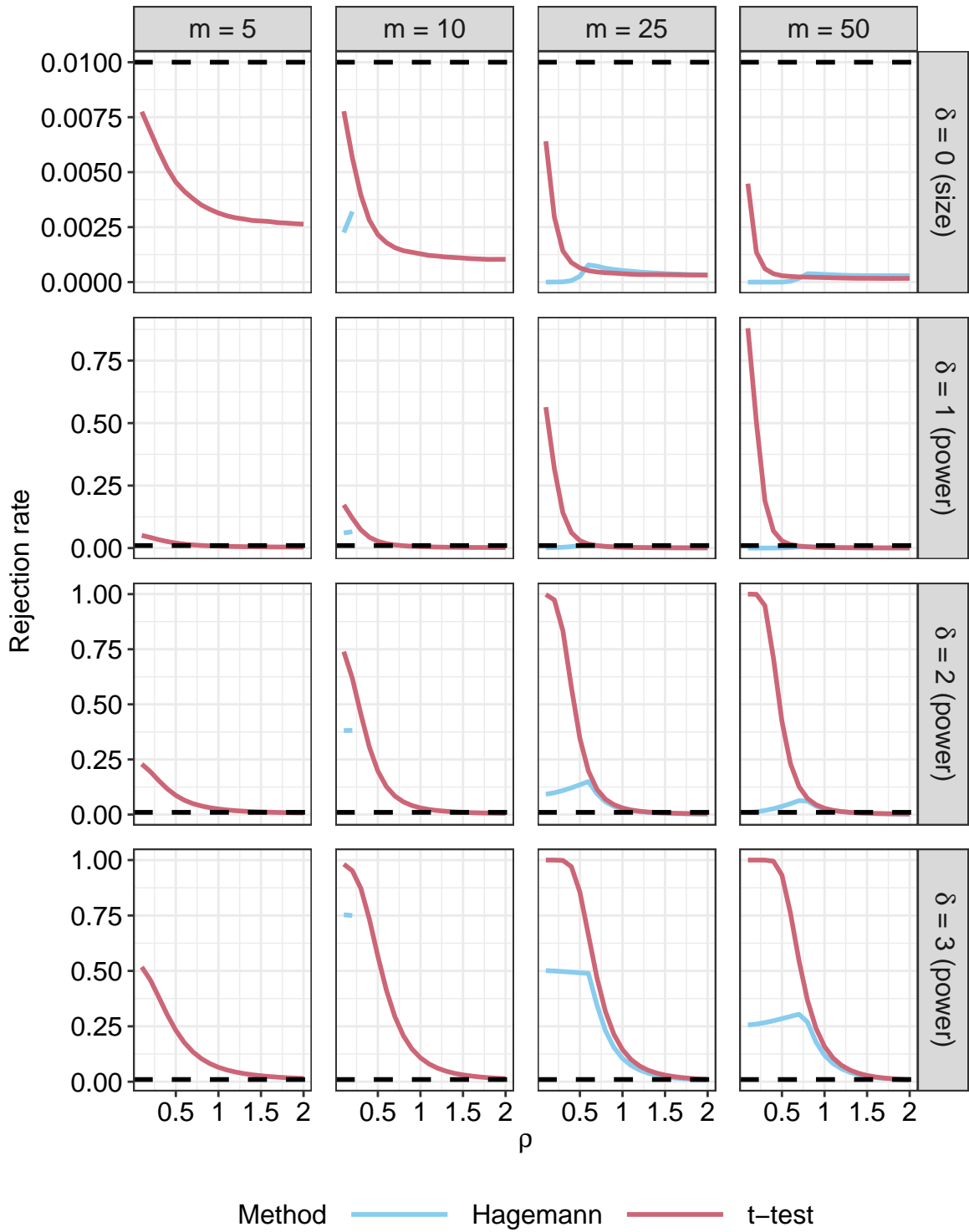


Figure S3: Probability of rejection against heterogeneity parameter ρ for various cluster size m and alternatives θ at $\alpha = 0.01$ for **DGP 2** of simulation design 1 with $k = 1$.

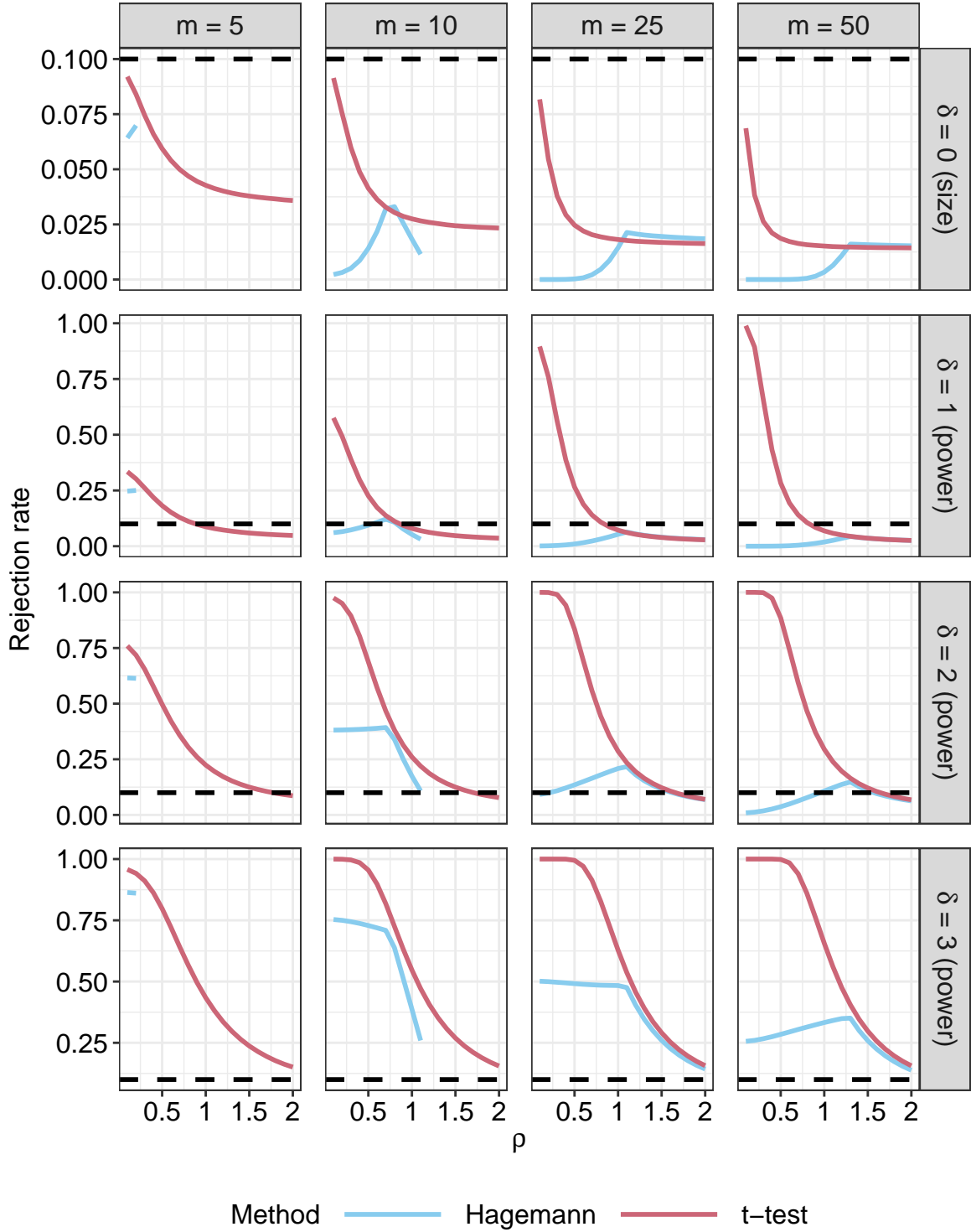


Figure S4: Probability of rejection against heterogeneity parameter ρ for various cluster size m and alternatives θ at $\alpha = 0.1$ for **DGP 2** of simulation design 1 with $k = 1$.

S3.1.2 Results for $k = 2$

This subsection reports the results for $k = 2$. Figures S5 to S7 report the results at the 1%, 5% and 10% levels respectively for **DGP 1**. Figures S8 to S10 report the results at the 1%, 5% and 10% levels respectively for **DGP 2**. The figures show that the t -test continue to perform favorably in other significance levels.

S3.2 Supplemental results for Section 6.2

Figures S11 to S13 report the results for DGPs 3 to 5 for various m and ρ at $\alpha = 0.05$.

S4 Supplemental tables: critical values for $k = 2$

Table S1 reports the critical values for different numbers of clusters m and heterogeneity parameters ρ for $\alpha = 0.01$ and $\alpha = 0.05$ when $k = 2$.

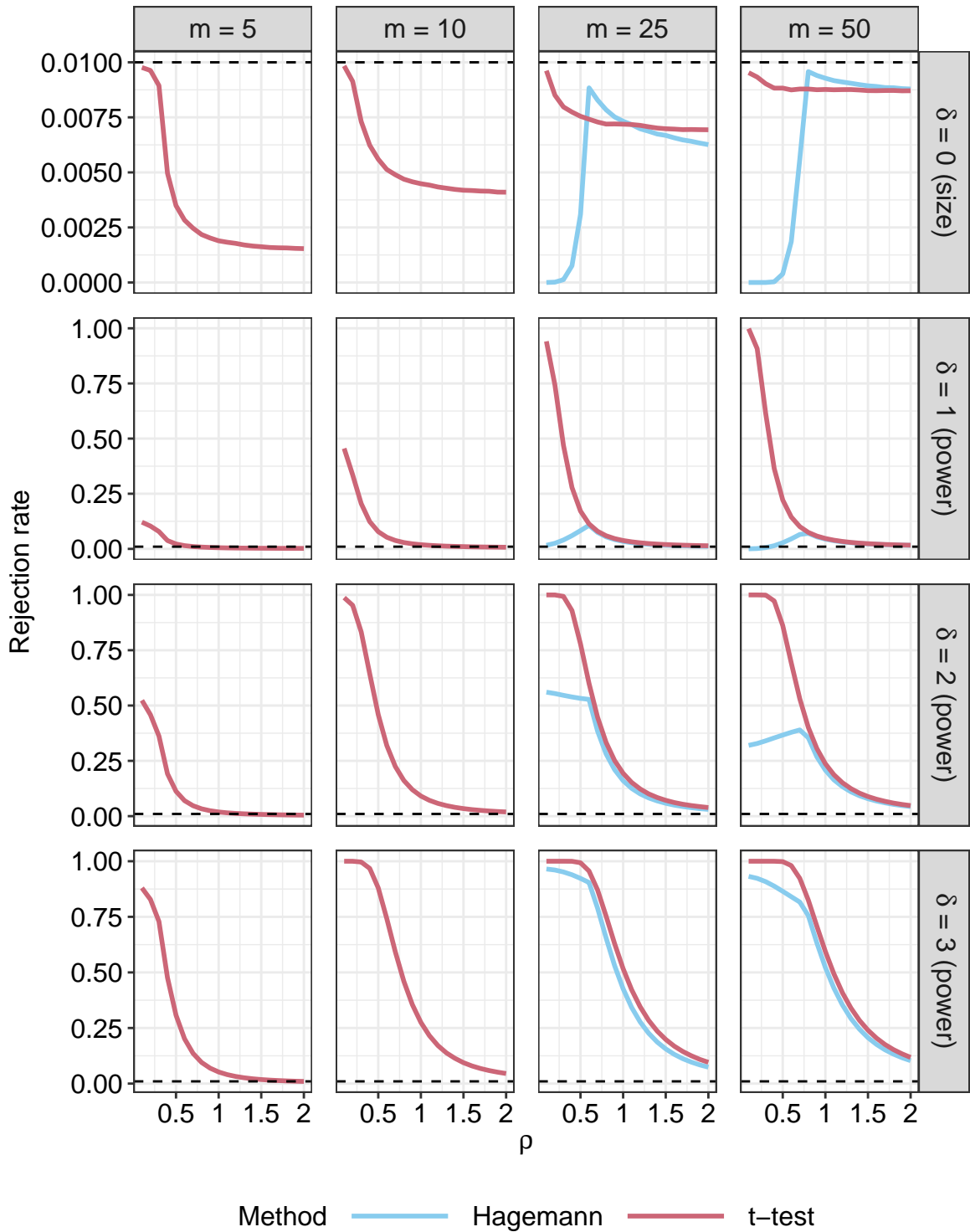


Figure S5: Probability of rejection against heterogeneity parameter ρ for various cluster size m and alternatives δ at $\alpha = 0.01$ for **DGP 1** of simulation design 1 with $k = 2$.

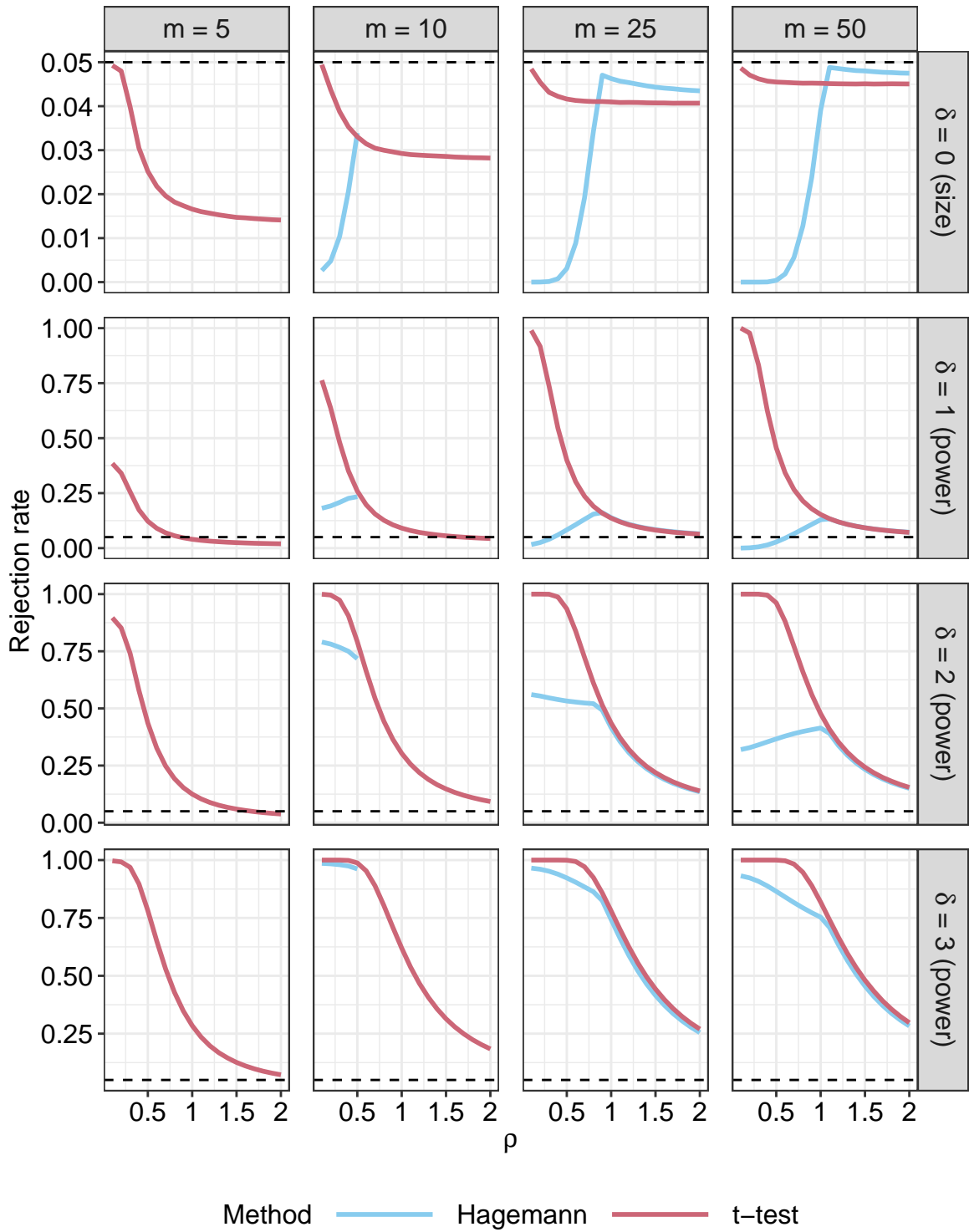


Figure S6: Probability of rejection against heterogeneity parameter ρ for various cluster size m and alternatives δ at $\alpha = 0.05$ for **DGP 1** of simulation design 1 with $k = 2$.

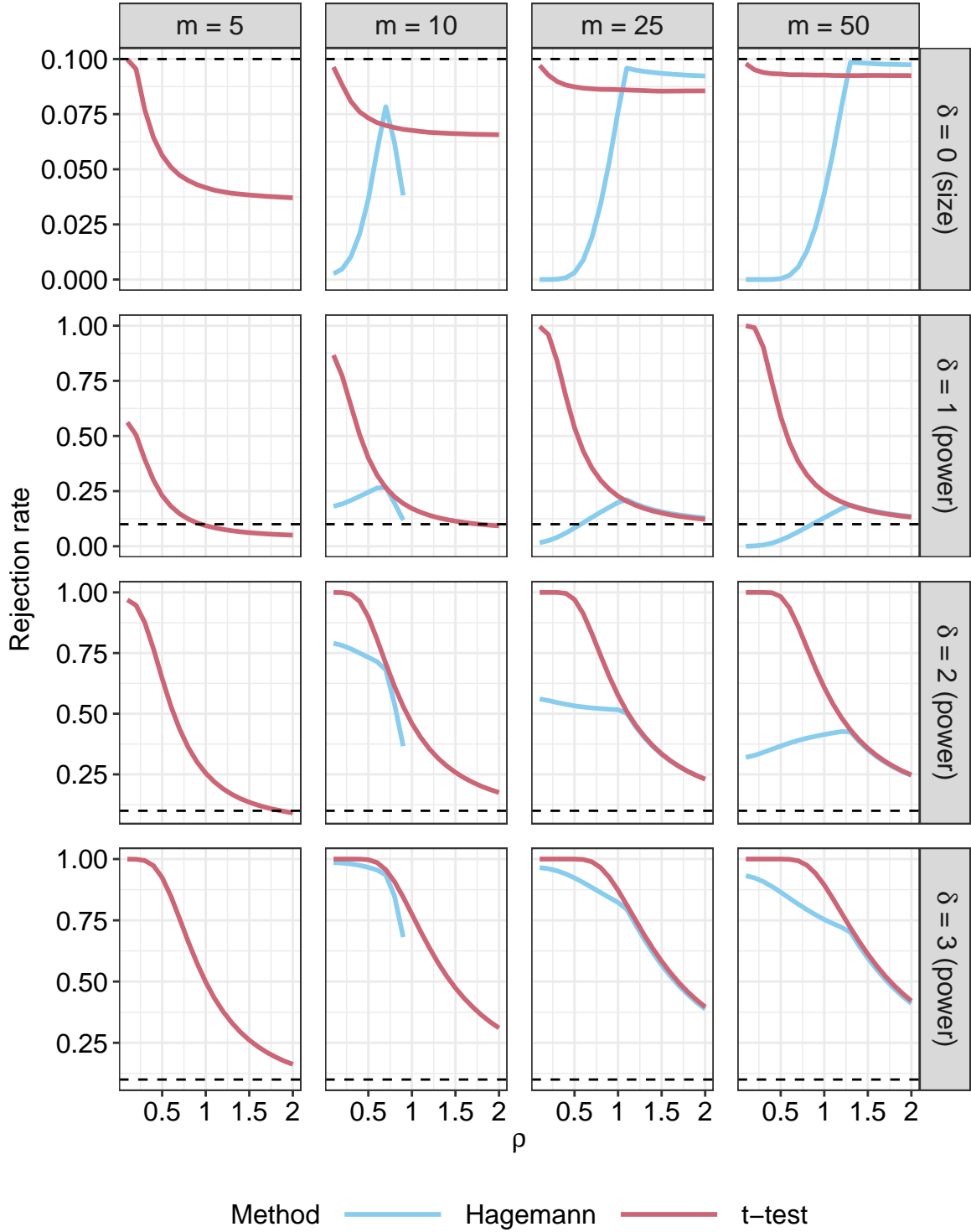


Figure S7: Probability of rejection against heterogeneity parameter ρ for various cluster size m and alternatives δ at $\alpha = 0.1$ for **DGP 1** of simulation design 1 with $k = 2$.

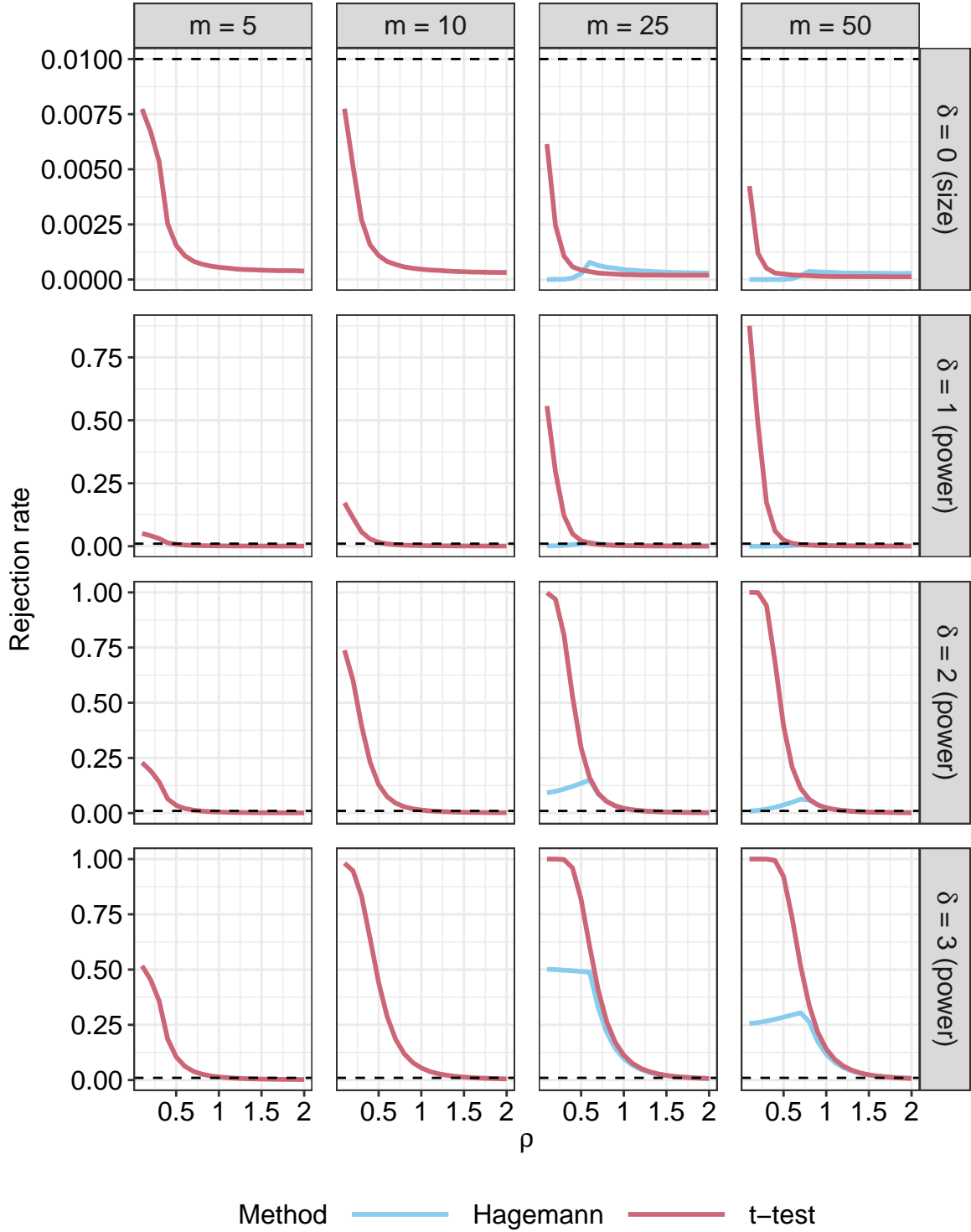


Figure S8: Probability of rejection against heterogeneity parameter ρ for various cluster size m and alternatives δ at $\alpha = 0.01$ for **DGP 2** of simulation design 1 with $k = 2$.

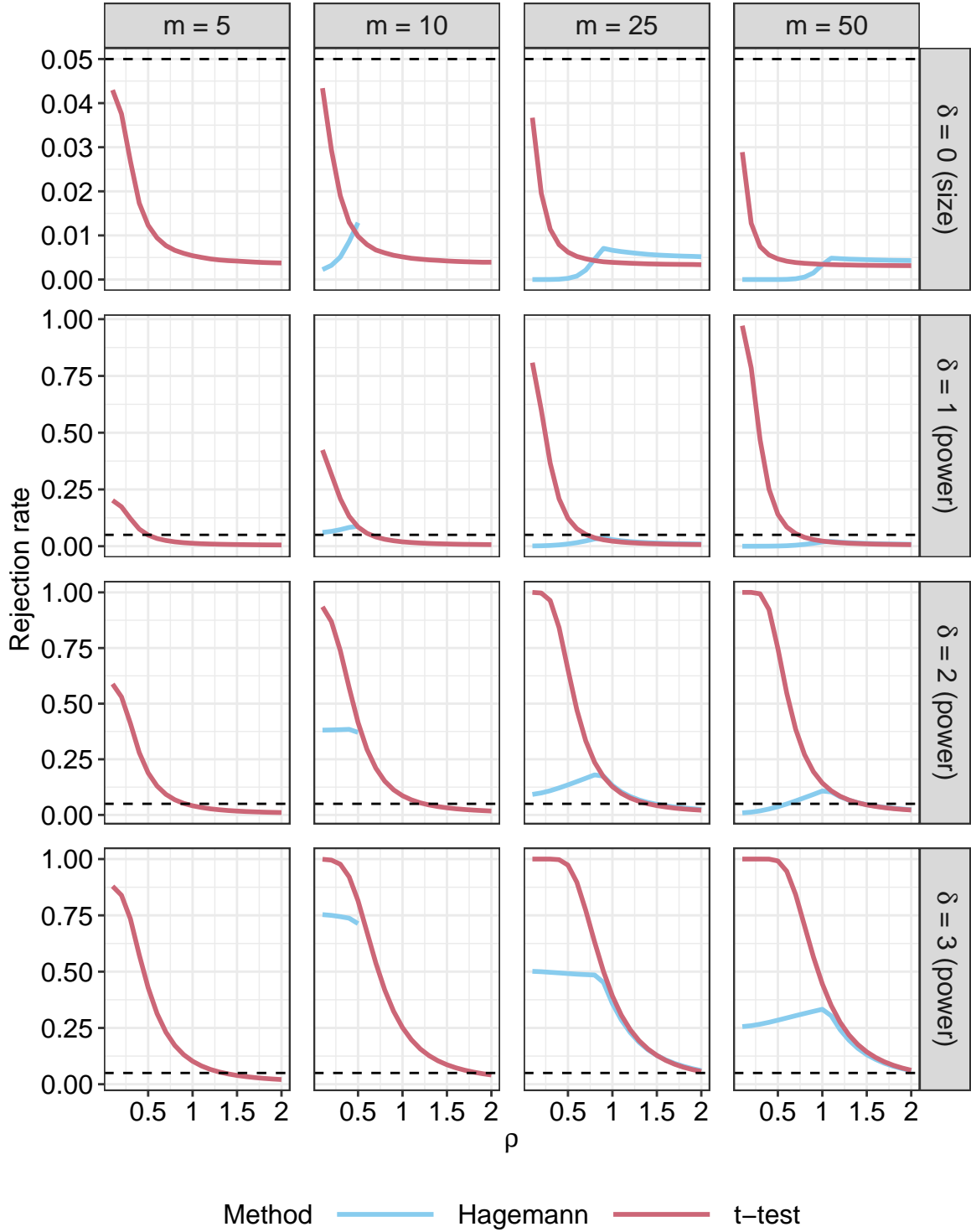


Figure S9: Probability of rejection against heterogeneity parameter ρ for various cluster size m and alternatives δ at $\alpha = 0.05$ for **DGP 2** of simulation design 1 with $k = 2$.

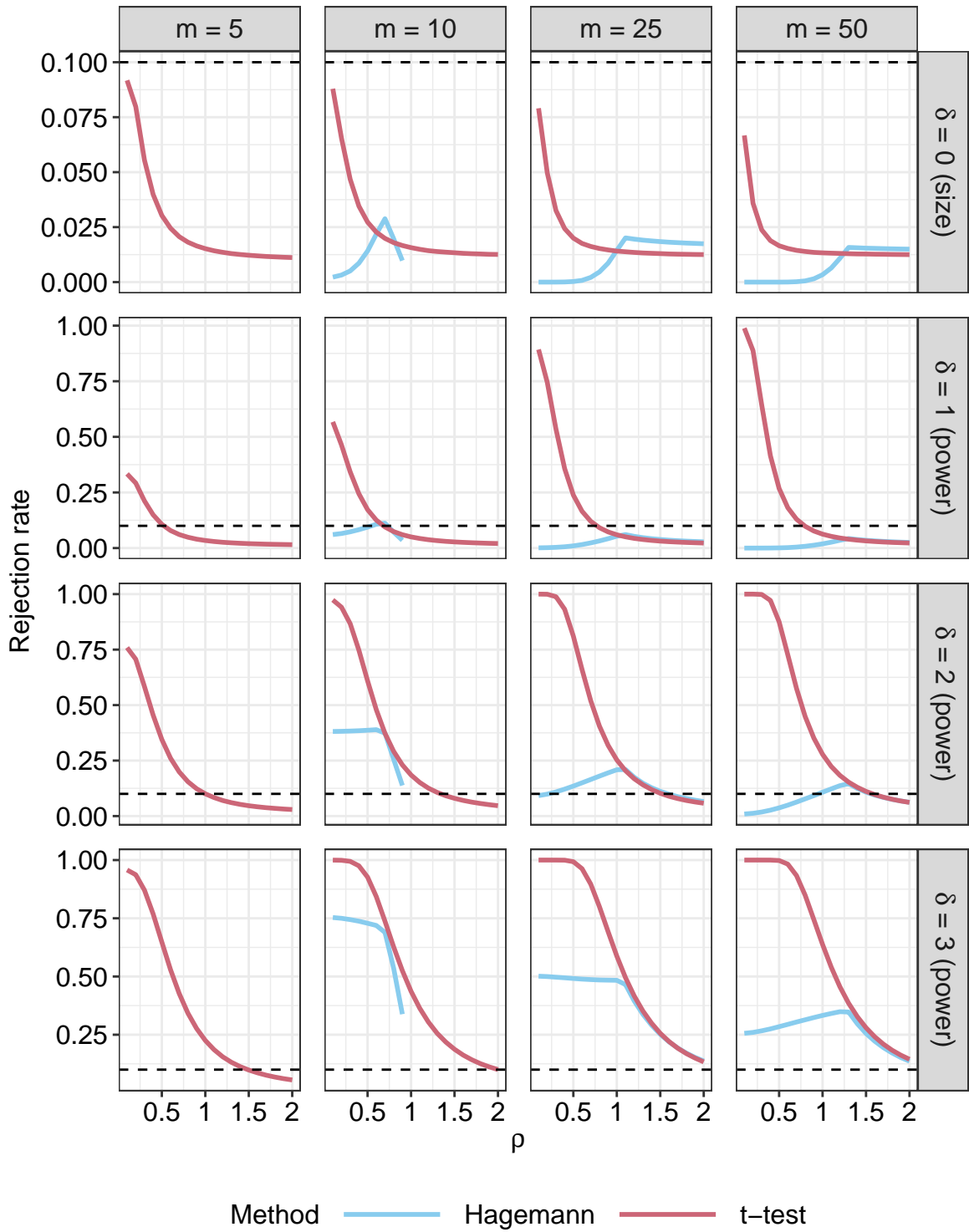


Figure S10: Probability of rejection against heterogeneity parameter ρ for various cluster size m and alternatives δ at $\alpha = 0.1$ for **DGP 2** of simulation design 1 with $k = 2$.

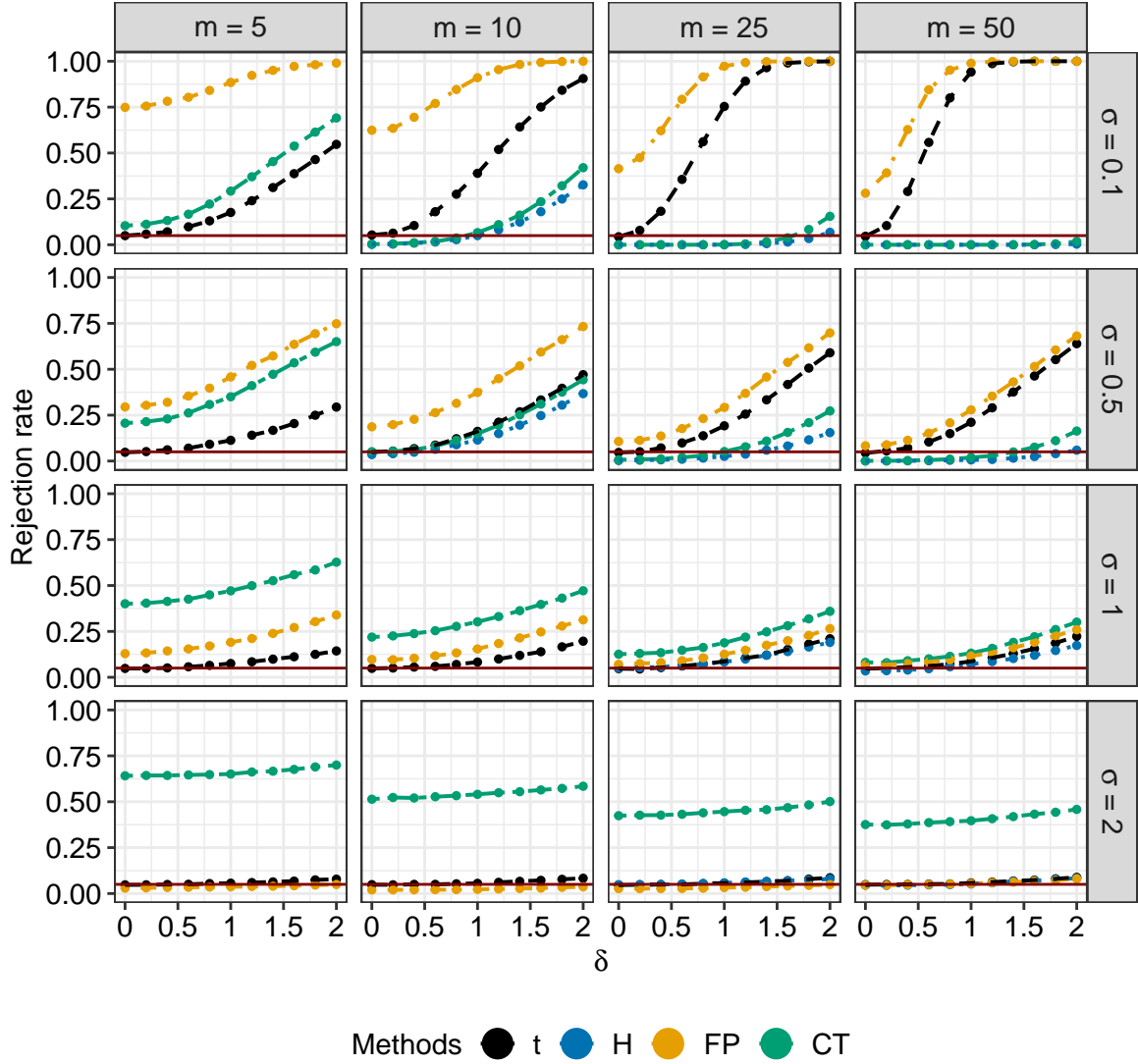


Figure S11: Simulation results for **DGP 3** of simulation design 2 at $\alpha = 0.05$.

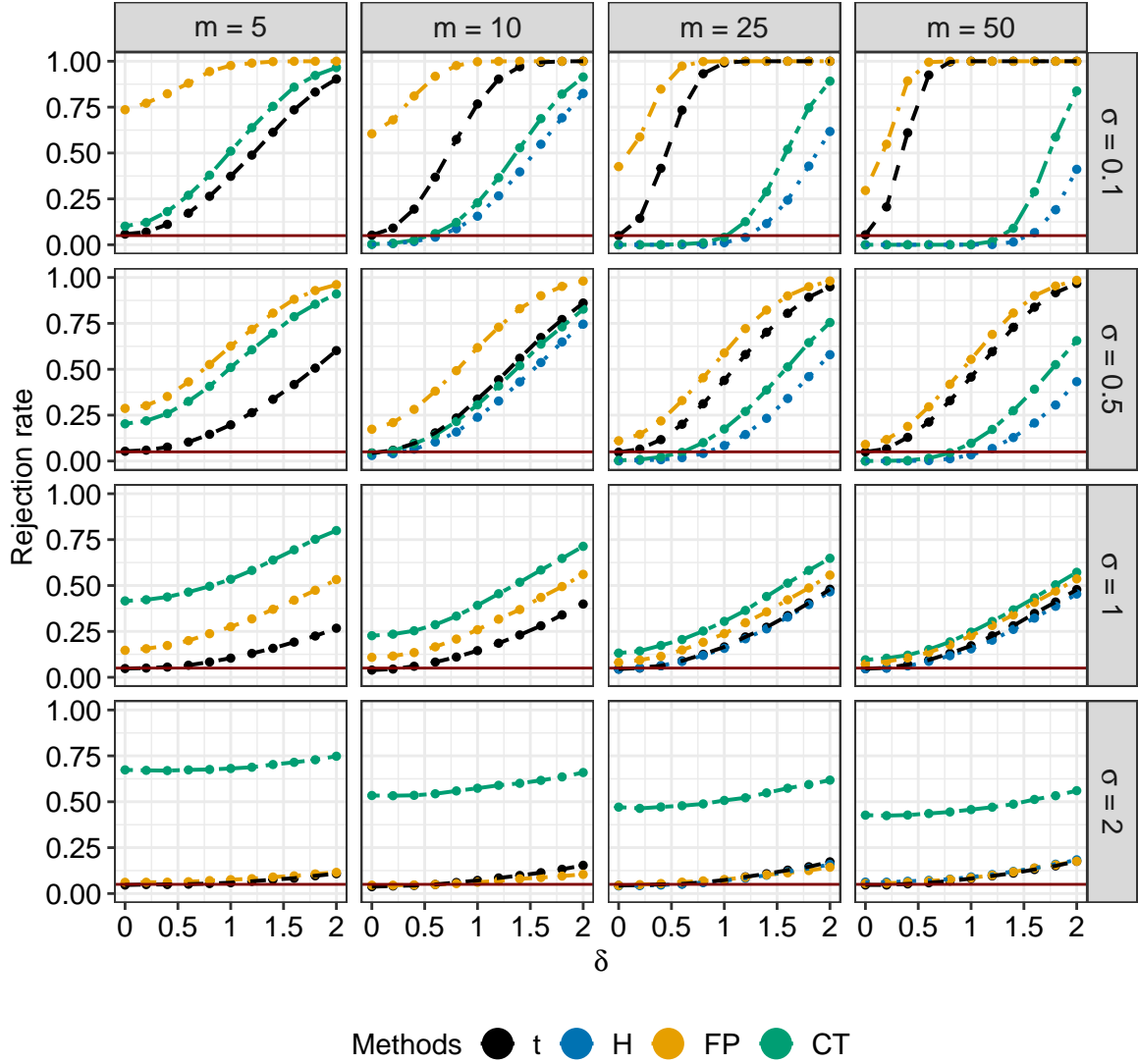


Figure S12: Simulation results for **DGP 4** of simulation design 2 at $\alpha = 0.05$.

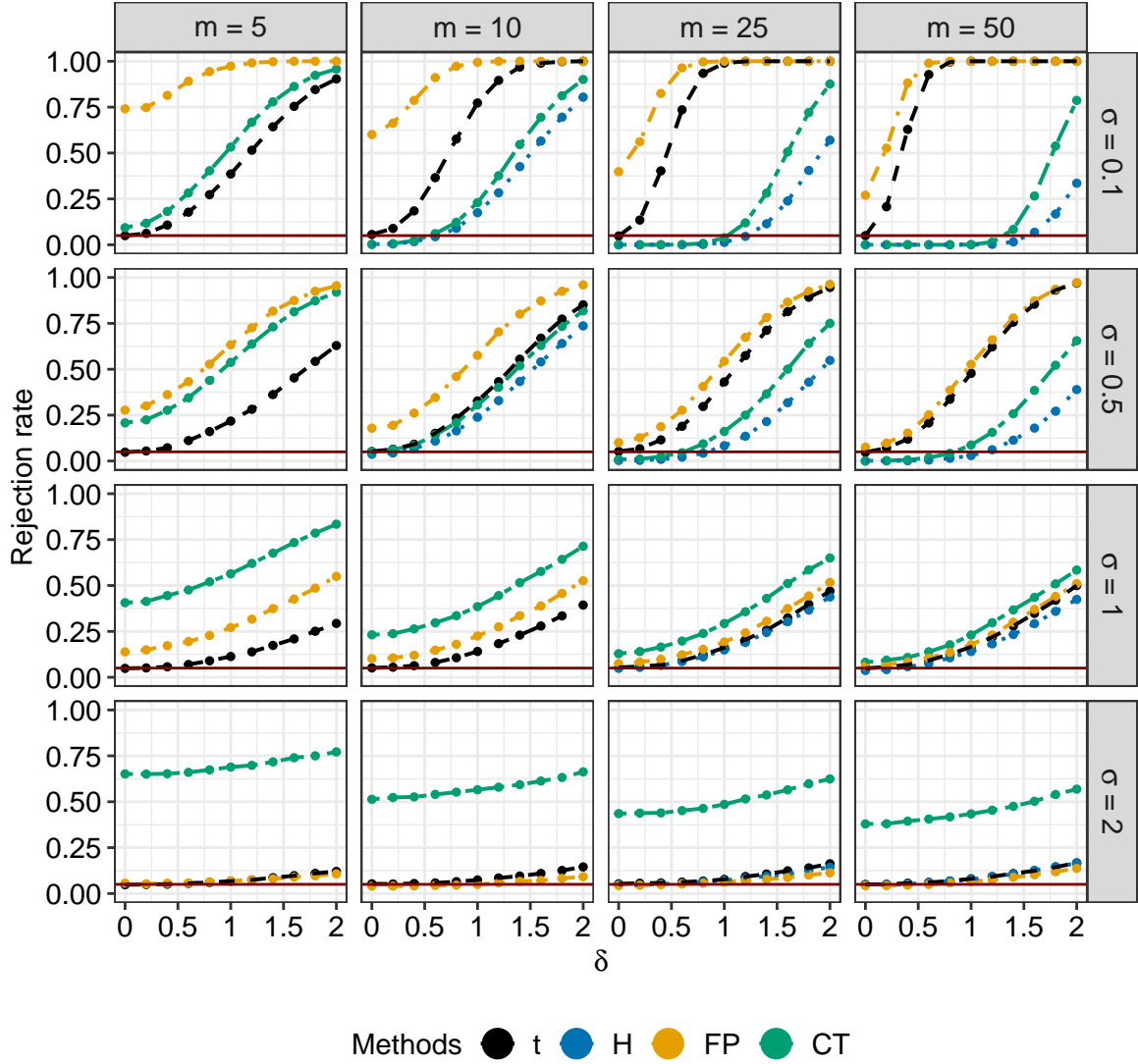


Figure S13: Simulation results for **DGP 5** of simulation design 2 at $\alpha = 0.05$.

Table S1: Critical values for different values of α , m and ρ for $k = 2$.

$\rho \backslash m$	$\alpha = 0.01$						$\alpha = 0.05$					
	5	10	15	20	25	50	5	10	15	20	25	50
0.2	2.260	1.224	0.984	0.869	0.800	0.661	1.365	0.860	0.711	0.636	0.590	0.496
0.4	2.820	1.722	1.460	1.341	1.274	1.148	1.729	1.200	1.051	0.981	0.940	0.861
0.6	3.652	2.328	2.017	1.882	1.807	1.669	2.264	1.616	1.450	1.375	1.332	1.251
0.8	4.668	2.977	2.604	2.445	2.358	2.200	2.849	2.061	1.870	1.786	1.738	1.649
1.0	5.724	3.644	3.204	3.019	2.918	2.736	3.459	2.521	2.301	2.205	2.151	2.051
1.2	6.794	4.323	3.811	3.598	3.482	3.273	4.082	2.988	2.737	2.627	2.566	2.454
1.4	7.874	5.007	4.423	4.180	4.049	3.812	4.713	3.460	3.176	3.052	2.984	2.858
1.6	8.960	5.696	5.037	4.765	4.617	4.351	5.350	3.935	3.616	3.479	3.403	3.262
1.8	10.049	6.387	5.653	5.350	5.186	4.892	5.990	4.411	4.058	3.906	3.822	3.667
2.0	11.142	7.080	6.270	5.936	5.756	5.432	6.633	4.889	4.501	4.334	4.242	4.072
2.2	12.236	7.775	6.888	6.524	6.326	5.973	7.278	5.368	4.945	4.763	4.662	4.477
2.4	13.332	8.470	7.507	7.111	6.897	6.513	7.924	5.848	5.389	5.192	5.083	4.883
2.6	14.429	9.166	8.126	7.699	7.468	7.054	8.572	6.329	5.834	5.621	5.504	5.288
2.8	15.527	9.864	8.746	8.288	8.039	7.596	9.220	6.810	6.279	6.051	5.925	5.694
3.0	16.626	10.561	9.366	8.876	8.611	8.137	9.869	7.291	6.724	6.481	6.346	6.100
3.2	17.726	11.259	9.987	9.465	9.183	8.678	10.519	7.773	7.169	6.911	6.767	6.505
3.4	18.826	11.957	10.607	10.054	9.754	9.220	11.169	8.254	7.614	7.340	7.189	6.911
3.6	19.926	12.656	11.228	10.643	10.326	9.761	11.819	8.736	8.060	7.771	7.610	7.317
3.8	21.027	13.355	11.849	11.232	10.898	10.302	12.470	9.219	8.506	8.201	8.032	7.723
4.0	22.128	14.054	12.470	11.821	11.471	10.844	13.121	9.701	8.952	8.631	8.454	8.129
4.2	23.229	14.753	13.092	12.411	12.043	11.386	13.772	10.184	9.398	9.061	8.875	8.535
4.4	24.331	15.452	13.713	13.000	12.615	11.927	14.424	10.666	9.844	9.492	9.297	8.941
4.6	25.432	16.152	14.334	13.590	13.187	12.469	15.075	11.149	10.290	9.922	9.719	9.347
4.8	26.534	16.851	14.956	14.180	13.760	13.010	15.727	11.632	10.736	10.353	10.141	9.753
5.0	27.636	17.551	15.578	14.769	14.332	13.552	16.379	12.115	11.182	10.783	10.562	10.159