

# Supplementary materials for *Small sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models*

July 18, 2016

## Contents

<b>S1 Proof of Theorem 1</b>	<b>2</b>
<b>S2 Proof of Theorem 2</b>	<b>3</b>
<b>S3 Details of simulation study</b>	<b>4</b>
<b>S4 Additional simulation results</b>	<b>6</b>
S4.1 Rejection rates of AHT and standard tests . . . . .	6
S4.2 Rejection rates of AHT and standard tests by study design . . . . .	9
S4.3 Rejection rates of AHT test using CR1 or CR2, with and without accounting for absorption .	21
S4.4 Rejection rates of AHT test by degree of working model misspecification . . . . .	25

## S1 Proof of Theorem 1

The Moore-Penrose inverse of  $\mathbf{B}_i$  can be computed from its eigen-decomposition. Let  $b \leq n_i$  denote the rank of  $\mathbf{B}_i$ . Let  $\mathbf{A}$  be the  $b \times b$  diagonal matrix of the positive eigenvalues of  $\mathbf{B}_i$  and  $\mathbf{V}$  be the  $n_i \times b$  matrix of corresponding eigen-vectors, so that  $\mathbf{B}_i = \mathbf{V}\mathbf{A}\mathbf{V}'$ . Then  $\mathbf{B}_i^+ = \mathbf{V}\mathbf{A}^{-1}\mathbf{V}'$  and  $\mathbf{B}_i^{+1/2} = \mathbf{V}\mathbf{A}^{-1/2}\mathbf{V}'$ . Now, observe that

$$\begin{aligned} \ddot{\mathbf{R}}_i' \mathbf{W}_i \mathbf{A}_i (\mathbf{I} - \mathbf{H}_{\mathbf{X}})_i \Phi (\mathbf{I} - \mathbf{H}_{\mathbf{X}})_i' \mathbf{A}_i' \mathbf{W}_i \ddot{\mathbf{R}}_i &= \ddot{\mathbf{R}}_i' \mathbf{W}_i \mathbf{D}_i \mathbf{B}_i^{+1/2} \mathbf{B}_i \mathbf{B}_i^{+1/2} \mathbf{D}_i' \mathbf{W}_i \ddot{\mathbf{R}}_i \\ &= \ddot{\mathbf{R}}_i' \mathbf{W}_i \mathbf{D}_i \mathbf{V} \mathbf{V}' \mathbf{D}_i' \mathbf{W}_i \ddot{\mathbf{R}}_i. \end{aligned} \quad (1)$$

Because  $\mathbf{D}_i$ , and  $\Phi$  are positive definite and  $\mathbf{B}_i$  is symmetric, the eigen-vectors  $\mathbf{V}$  define an orthonormal basis for the column span of  $(\mathbf{I} - \mathbf{H}_{\mathbf{X}})_i$ . We now show that  $\ddot{\mathbf{U}}_i$  is in the column space of  $(\mathbf{I} - \mathbf{H}_{\mathbf{X}})_i$ . Let  $\mathbf{Z}_i$  be an  $n_i \times (r + s)$  matrix of zeros. Let  $\mathbf{Z}_k = -\ddot{\mathbf{U}}_k \mathbf{L}^{-1} \mathbf{M}_{\ddot{\mathbf{U}}}^{-1}$ , for  $k \neq i$  and take  $\mathbf{Z} = (\mathbf{Z}_1', \dots, \mathbf{Z}_m')'$ . Now observe that  $(\mathbf{I} - \mathbf{H}_{\mathbf{T}}) \mathbf{Z} = \mathbf{Z}$ . It follows that

$$\begin{aligned} (\mathbf{I} - \mathbf{H}_{\mathbf{X}})_i \mathbf{Z} &= (\mathbf{I} - \mathbf{H}_{\ddot{\mathbf{U}}})_i (\mathbf{I} - \mathbf{H}_{\mathbf{T}}) \mathbf{Z} = (\mathbf{I} - \mathbf{H}_{\ddot{\mathbf{U}}})_i \mathbf{Z} \\ &= \mathbf{Z}_i - \ddot{\mathbf{U}}_i \mathbf{M}_{\ddot{\mathbf{U}}} \sum_{k=1}^m \ddot{\mathbf{U}}_k' \mathbf{W}_k \mathbf{Z}_k \\ &= \ddot{\mathbf{U}}_i \mathbf{M}_{\ddot{\mathbf{U}}} \left( \sum_{k \neq i} \ddot{\mathbf{U}}_k' \mathbf{W}_k \ddot{\mathbf{U}} \right) \mathbf{L}^{-1} \mathbf{M}_{\ddot{\mathbf{U}}}^{-1} = \ddot{\mathbf{U}}_i. \end{aligned}$$

Thus, there exists an  $N \times (r + s)$  matrix  $\mathbf{Z}$  such that  $(\mathbf{I} - \mathbf{H}_{\ddot{\mathbf{X}}})_i \mathbf{Z} = \ddot{\mathbf{U}}_i$ , i.e.,  $\ddot{\mathbf{U}}_i$  is in the column span of  $(\mathbf{I} - \mathbf{H}_{\mathbf{X}})_i$ . Because  $\mathbf{D}_i \mathbf{W}_i$  is positive definite and  $\ddot{\mathbf{R}}_i$  is a sub-matrix of  $\ddot{\mathbf{U}}_i$ ,  $\mathbf{D}_i \mathbf{W}_i \ddot{\mathbf{R}}_i$  is also in the column span of  $(\mathbf{I} - \mathbf{H}_{\mathbf{X}})_i$ . It follows that

$$\ddot{\mathbf{R}}_i' \mathbf{W}_i \mathbf{D}_i \mathbf{V} \mathbf{V}' \mathbf{D}_i' \mathbf{W}_i \ddot{\mathbf{R}}_i = \ddot{\mathbf{R}}_i' \mathbf{W}_i \Phi \mathbf{W}_i \ddot{\mathbf{R}}_i. \quad (2)$$

Substituting (2) into (1) demonstrates that  $\mathbf{A}_i$  satisfies the generalized BRL criterion (Eq. 6 of the main paper).

Under the working model, the residuals from cluster  $i$  have mean  $\mathbf{0}$  and variance

$$\text{Var}(\ddot{\mathbf{e}}_i) = (\mathbf{I} - \mathbf{H}_{\mathbf{X}})_i \Phi (\mathbf{I} - \mathbf{H}_{\mathbf{X}})_i',$$

It follows that

$$\begin{aligned} \mathbb{E}(\mathbf{V}^{CR2}) &= \mathbf{M}_{\ddot{\mathbf{R}}} \left[ \sum_{i=1}^m \ddot{\mathbf{R}}_i' \mathbf{W}_i \mathbf{A}_i (\mathbf{I} - \mathbf{H}_{\mathbf{X}})_i \Phi (\mathbf{I} - \mathbf{H}_{\mathbf{X}})_i' \mathbf{A}_i \mathbf{W}_i \ddot{\mathbf{R}}_i \right] \mathbf{M}_{\ddot{\mathbf{R}}} \\ &= \mathbf{M}_{\ddot{\mathbf{R}}} \left[ \sum_{i=1}^m \ddot{\mathbf{R}}_i' \mathbf{W}_i \Phi \mathbf{W}_i \ddot{\mathbf{R}}_i \right] \mathbf{M}_{\ddot{\mathbf{R}}} \\ &= \text{Var}(\hat{\beta}) \end{aligned}$$

## S2 Proof of Theorem 2

From the fact that  $\ddot{\mathbf{U}}_i' \mathbf{W}_i \mathbf{T}_i = \mathbf{0}$  for  $i = 1, \dots, m$ , it follows that

$$\begin{aligned} \mathbf{B}_i &= \mathbf{D}_i (\mathbf{I} - \mathbf{H}_{\ddot{\mathbf{U}}})_i (\mathbf{I} - \mathbf{H}_{\mathbf{T}}) \Phi (\mathbf{I} - \mathbf{H}_{\mathbf{T}})' (\mathbf{I} - \mathbf{H}_{\ddot{\mathbf{U}}})_i' \mathbf{D}_i' \\ &= \mathbf{D}_i (\mathbf{I} - \mathbf{H}_{\ddot{\mathbf{U}}} - \mathbf{H}_{\mathbf{T}})_i \Phi (\mathbf{I} - \mathbf{H}_{\ddot{\mathbf{U}}} - \mathbf{H}_{\mathbf{T}})'_i \mathbf{D}_i' \\ &= \mathbf{D}_i \left( \Phi_i - \ddot{\mathbf{U}}_i \mathbf{M}_{\ddot{\mathbf{U}}} \ddot{\mathbf{U}}_i' - \mathbf{T}_i \mathbf{M}_{\mathbf{T}} \mathbf{T}_i' \right) \mathbf{D}_i' \end{aligned}$$

and

$$\mathbf{B}_i^+ = (\mathbf{D}_i')^{-1} \left( \Phi_i - \ddot{\mathbf{U}}_i \mathbf{M}_{\ddot{\mathbf{U}}} \ddot{\mathbf{U}}_i' - \mathbf{T}_i \mathbf{M}_{\mathbf{T}} \mathbf{T}_i' \right)^+ \mathbf{D}_i^{-1}. \quad (3)$$

Let  $\Psi_i = \left( \Phi_i - \ddot{\mathbf{U}}_i \mathbf{M}_{\ddot{\mathbf{U}}} \ddot{\mathbf{U}}_i' \right)^+$ . Using a generalized Woodbury identity (Henderson and Searle, 1981),

$$\Psi_i = \mathbf{W}_i + \mathbf{W}_i \ddot{\mathbf{U}}_i \mathbf{M}_{\ddot{\mathbf{U}}} \left( \mathbf{M}_{\ddot{\mathbf{U}}} - \mathbf{M}_{\ddot{\mathbf{U}}} \ddot{\mathbf{U}}_i' \mathbf{W}_i \ddot{\mathbf{U}}_i \mathbf{M}_{\ddot{\mathbf{U}}} \right)^+ \mathbf{M}_{\ddot{\mathbf{U}}} \ddot{\mathbf{U}}_i' \mathbf{W}_i.$$

It follows that  $\Psi_i \mathbf{T}_i = \mathbf{W}_i \mathbf{T}_i$ . Another application of the generalized Woodbury identity gives

$$\begin{aligned} \left( \Phi_i - \ddot{\mathbf{U}}_i \mathbf{M}_{\ddot{\mathbf{U}}} \ddot{\mathbf{U}}_i' - \mathbf{T}_i \mathbf{M}_{\mathbf{T}} \mathbf{T}_i' \right)^+ &= \Psi_i + \Psi_i \mathbf{T}_i \mathbf{M}_{\mathbf{T}} (\mathbf{M}_{\mathbf{T}} - \mathbf{M}_{\mathbf{T}} \mathbf{T}_i' \Psi_i \mathbf{T}_i \mathbf{M}_{\mathbf{T}})^+ \mathbf{M}_{\mathbf{T}} \mathbf{T}_i' \Psi_i \\ &= \Psi_i + \mathbf{W}_i \mathbf{T}_i \mathbf{M}_{\mathbf{T}} (\mathbf{M}_{\mathbf{T}} - \mathbf{M}_{\mathbf{T}} \mathbf{T}_i' \mathbf{W}_i \mathbf{T}_i \mathbf{M}_{\mathbf{T}})^+ \mathbf{M}_{\mathbf{T}} \mathbf{T}_i' \mathbf{W}_i \\ &= \Psi_i. \end{aligned}$$

The last equality follows from the fact that

$$\mathbf{T}_i \mathbf{M}_{\mathbf{T}} (\mathbf{M}_{\mathbf{T}} - \mathbf{M}_{\mathbf{T}} \mathbf{T}_i' \mathbf{W}_i \mathbf{T}_i \mathbf{M}_{\mathbf{T}})^- \mathbf{M}_{\mathbf{T}} \mathbf{T}_i' = \mathbf{0}$$

because the fixed effects are nested within clusters. Substituting into (3), we then have that  $\mathbf{B}_i^+ = (\mathbf{D}_i')^{-1} \Psi_i \mathbf{D}_i^{-1}$ . But

$$\tilde{\mathbf{B}}_i = \mathbf{D}_i (\mathbf{I} - \mathbf{H}_{\ddot{\mathbf{U}}})_i \Phi (\mathbf{I} - \mathbf{H}_{\ddot{\mathbf{U}}})_i' \mathbf{D}_i' = \mathbf{D}_i \left( \Phi_i - \ddot{\mathbf{U}}_i \mathbf{M}_{\ddot{\mathbf{U}}} \ddot{\mathbf{U}}_i' \right) \mathbf{D}_i' = \mathbf{D}_i \Psi_i^+ \mathbf{D}_i',$$

and so  $\mathbf{B}_i^+ = \tilde{\mathbf{B}}_i^+$ . It follows that  $\mathbf{A}_i = \tilde{\mathbf{A}}_i$  for  $i = 1, \dots, m$ .

Table S1: Simulation design parameters

Parameter	levels
Design	RB, CR, DD
Balance	Balanced, Unbalanced
Outcome missingness	Complete data, 15% missing
Clusters ( $m$ )	15, 30, 50
Units per cluster ( $n$ )	12, 18, 30
Intra-class correlation ( $\tau^2$ )	.05, .15, .25
Treatment effect variability ( $\sigma_\delta^2$ )	.00, .04, .09
Correlation among outcomes ( $\rho$ )	.2, .8

### S3 Details of simulation study

This section provides further details regarding the design of the simulations reported in Section 4 of the main text. Table S1 summarizes the factors manipulated in the simulation.

The simulations examined six distinct study designs. Outcomes are measured for  $n$  units (which may be individuals, as in a cluster-randomized or block-randomized design, or time-points, as in a difference-in-differences panel) in each of  $m$  clusters under one of three treatment conditions. Suppose that there are  $G$  sets of clusters, each of size  $m_g$ , where the clusters in each set have a distinct configuration of treatment assignments. Let  $n_{ghi}$  denote the number of units at which cluster  $i$  in configuration  $g$  is observed under condition  $h$ , for  $i = 1, \dots, m$ ,  $g = 1, \dots, G$ , and  $h = 1, 2, 3$ . Table S2 summarizes the cluster-level sample sizes and unit-level patterns of treatment allocation for each of the six designs. The simulated designs included the following:

1. A balanced, block-randomized design, with an un-equal allocation within each block. In the balanced design, the treatment allocation is identical for each block, so  $G = 1$ .
2. An unbalanced, block-randomized design, with two different patterns of treatment allocation ( $G = 2$ ).
3. A balanced, cluster-randomized design, in which units are nested within clusters and an equal number of clusters are assigned to each treatment condition.
4. An unbalanced, cluster-randomized design, in which units are nested within clusters but the number of clusters assigned to each condition is not equal.
5. A balanced difference-in-differences design with two patterns of treatment allocation ( $G = 2$ ), in which half of the clusters are observed under the first treatment condition only and the remaining half are observed under all three conditions.
6. An unbalanced difference-in-differences design, again with two patterns of treatment allocation ( $G = 2$ ), but where 2/3 of the clusters are observed under the first treatment condition only and the remaining 1/3 of clusters are observed under all three conditions.

Table S2: Study designs used for simulation

Study design	Balance	Configuration	Clusters	Treatment allocation
Randomized Block	Balanced	1	$m_1 = m$	$n_{11i} = n/2, n_{12i} = n/3, n_{13i} = n/6$
Randomized Block	Unbalanced	1	$m_1 = m/2$	$n_{11i} = n/2, n_{12i} = n/3, n_{13i} = n/6$
		2	$m_2 = m/2$	$n_{21i} = n/3, n_{22i} = 5n/9, n_{23i} = n/9$
Cluster-Randomized	Balanced	1	$m_1 = m/3$	$n_{11i} = n$
		2	$m_2 = m/3$	$n_{22i} = n$
		3	$m_3 = m/3$	$n_{33i} = n$
Cluster-Randomized	Unbalanced	1	$m_1 = m/2$	$n_{11i} = n$
		2	$m_2 = 3m/10$	$n_{22i} = n$
		3	$m_3 = m/5$	$n_{33i} = n$
Difference-in-Differences	Balanced	1	$m_1 = m/2$	$n_{11i} = n$
		2	$m_2 = m/2$	$n_{21i} = n/2, n_{22i} = n/3, n_{23i} = n/6$
Difference-in-Differences	Unbalanced	1	$m_1 = 2m/3$	$n_{11i} = n$
		2	$m_2 = m/3$	$n_{21i} = n/2, n_{22i} = n/3, n_{23i} = n/6$

## S4 Additional simulation results

### S4.1 Rejection rates of AHT and standard tests

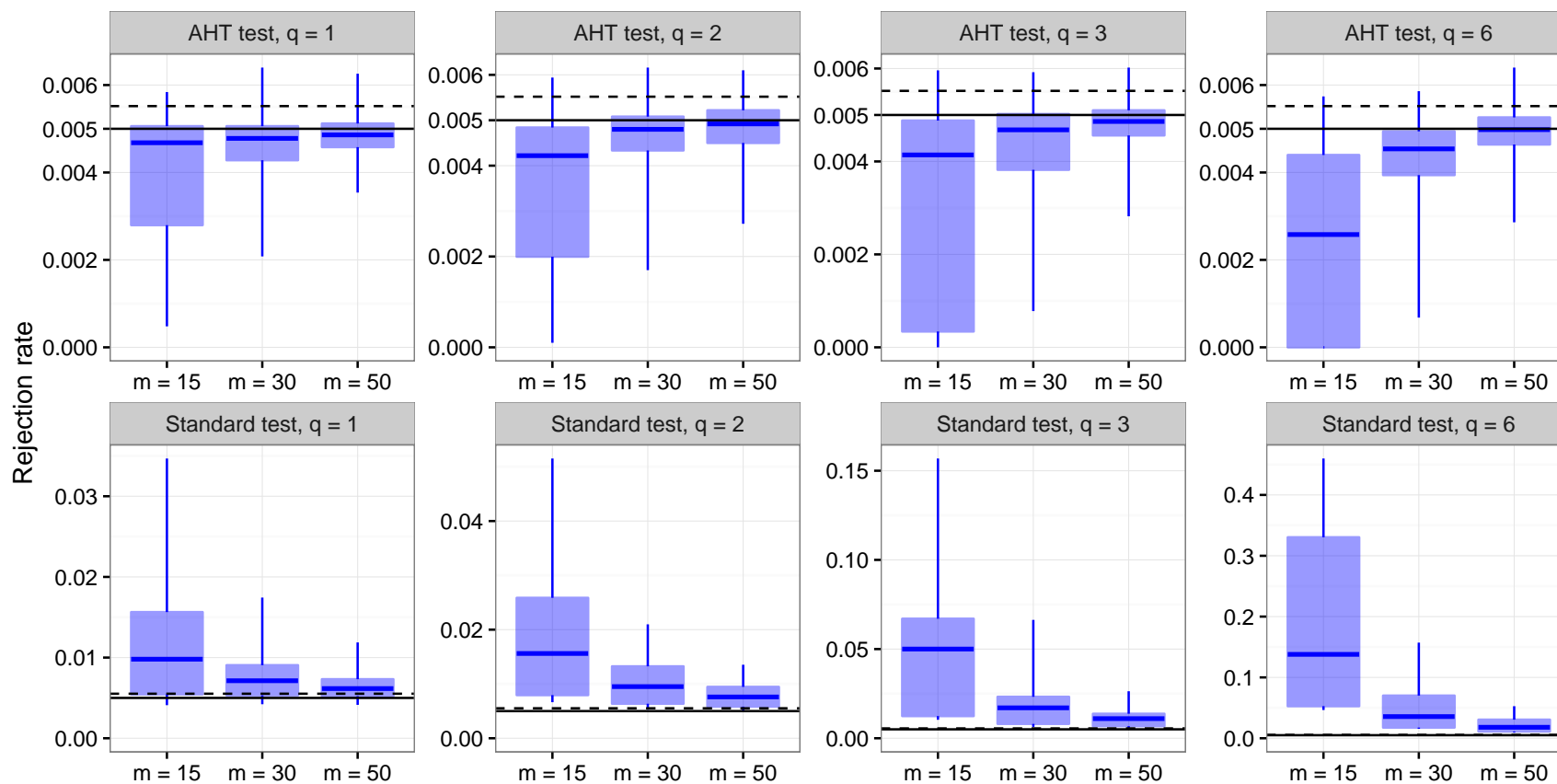


Figure S1: Rejection rates of AHT and standard tests for  $\alpha = .005$ , by dimension of hypothesis ( $q$ ) and sample size ( $m$ ).

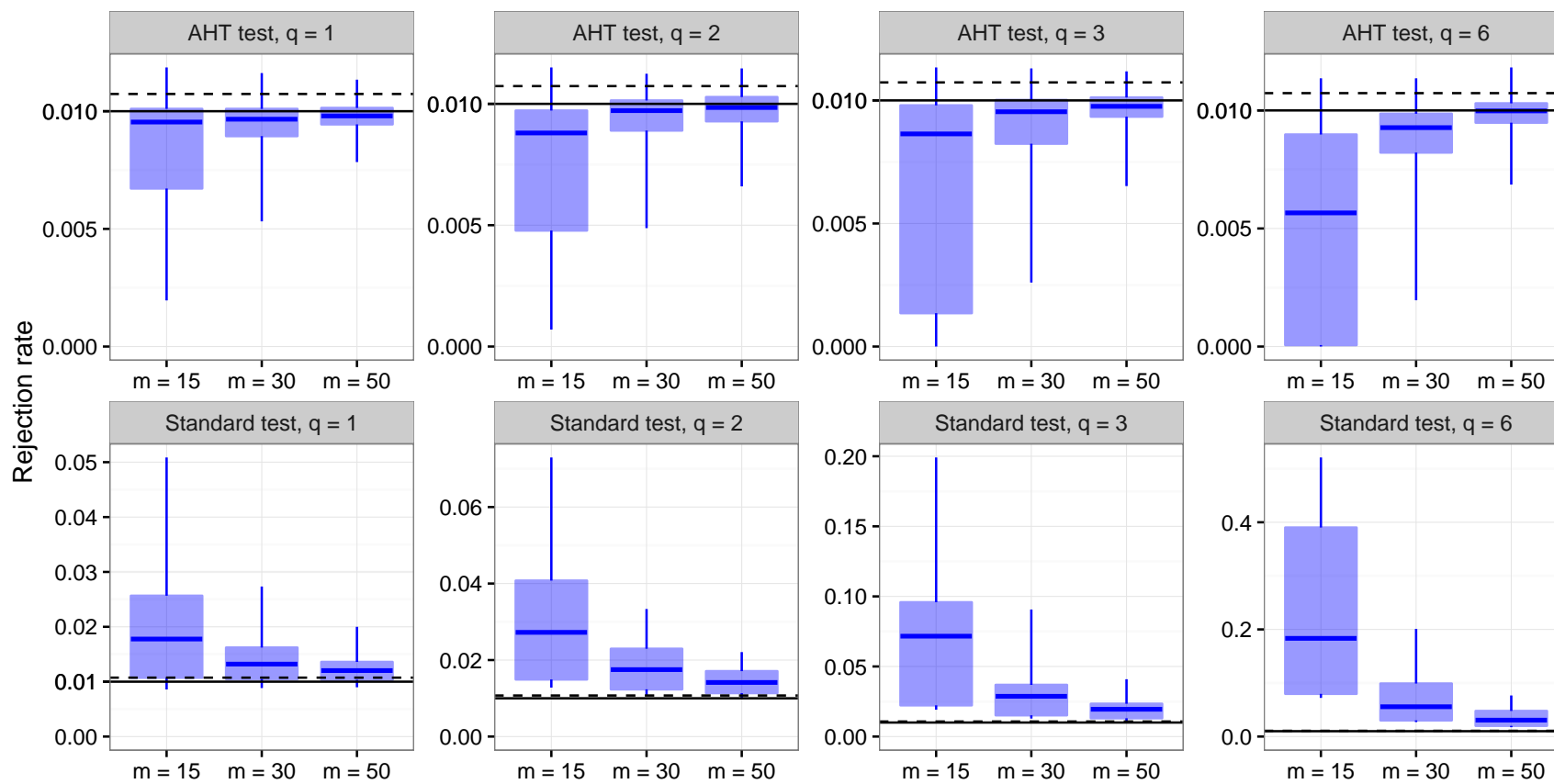


Figure S2: Rejection rates of AHT and standard tests for  $\alpha = .01$ , by dimension of hypothesis ( $q$ ) and sample size ( $m$ ).

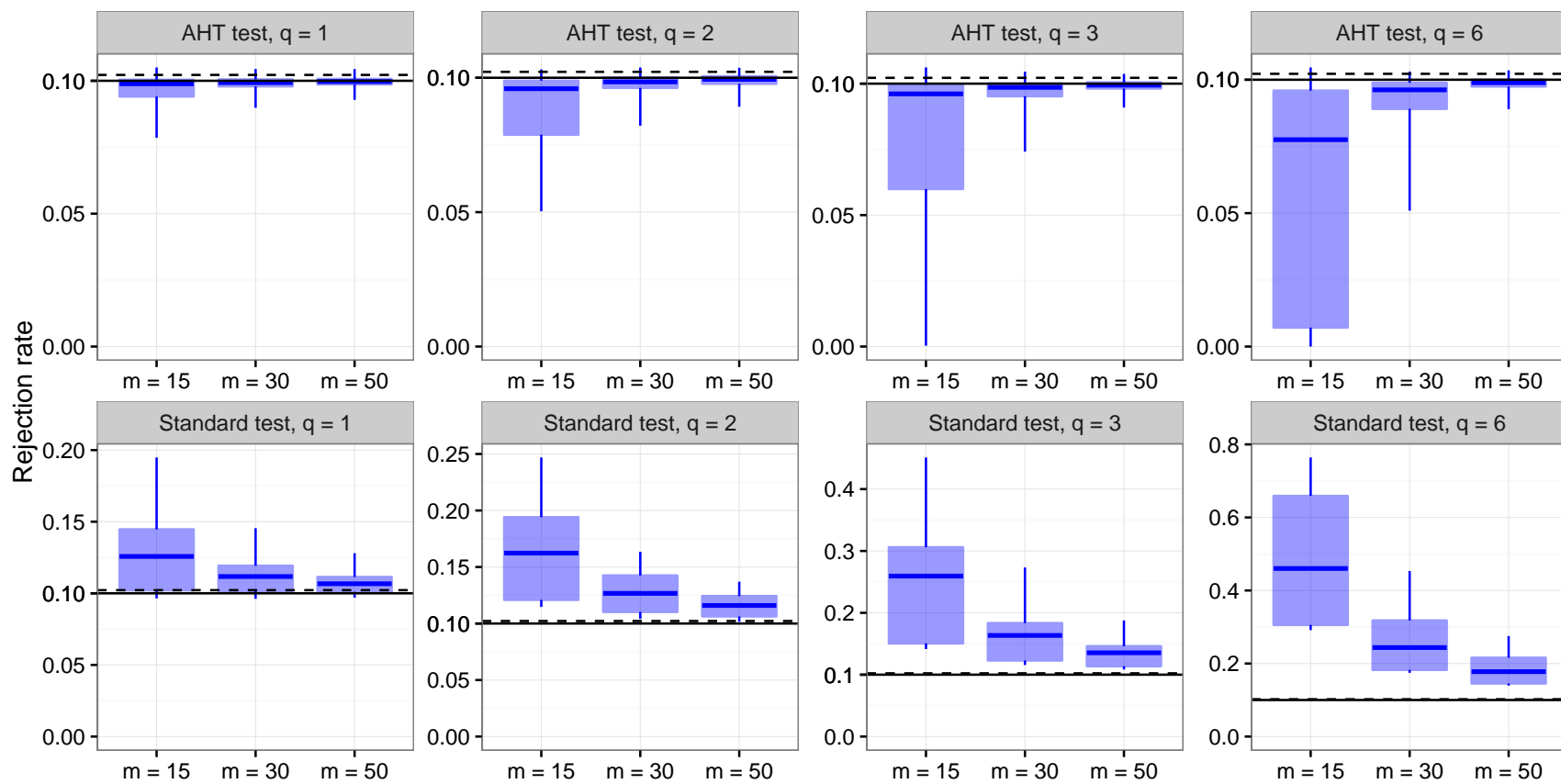


Figure S3: Rejection rates of AHT and standard tests for  $\alpha = .10$ , by dimension of hypothesis ( $q$ ) and sample size ( $m$ ).



## S4.2 Rejection rates of AHT and standard tests by study design

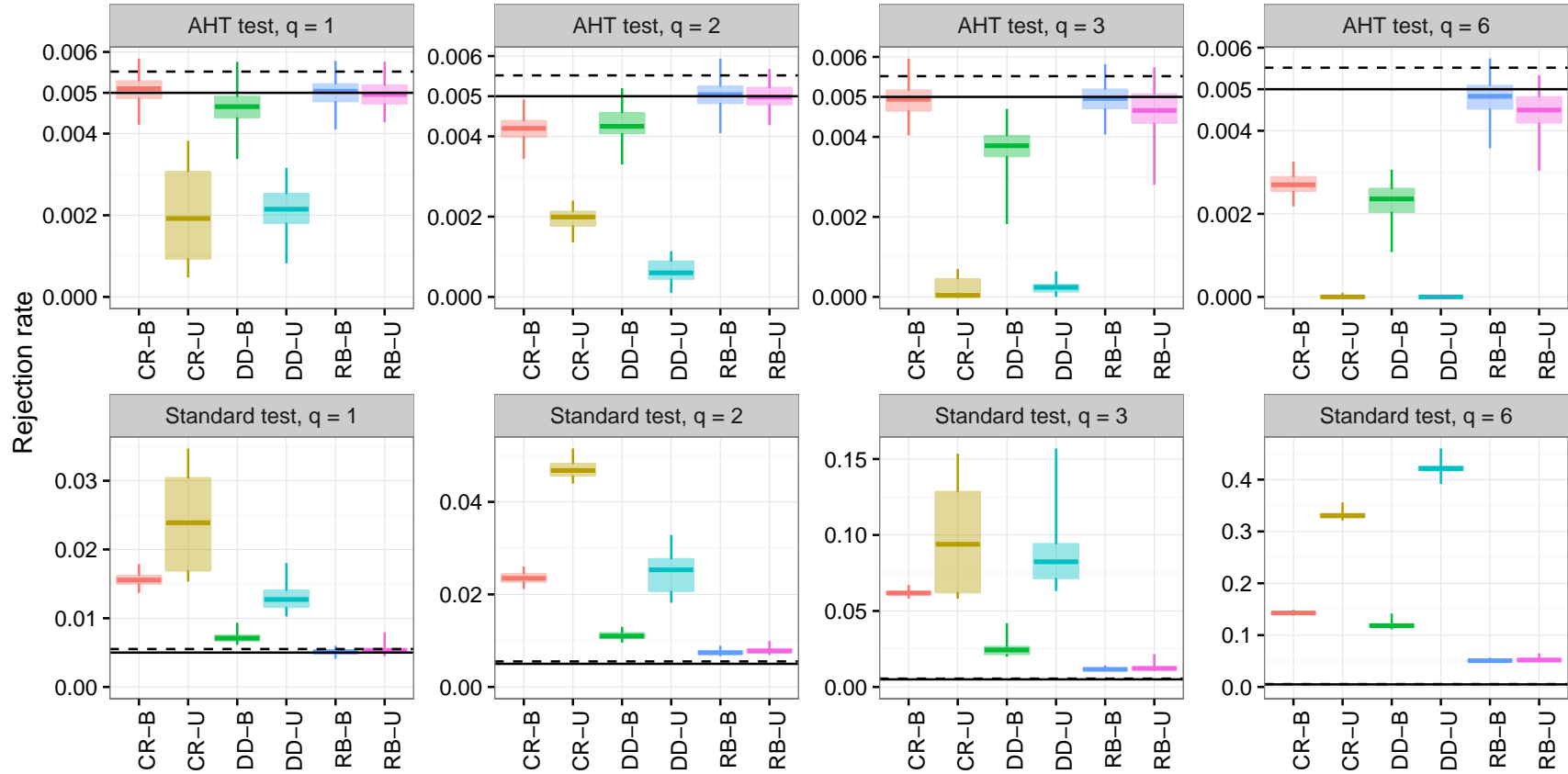


Figure S4: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis ( $q$ ) for  $\alpha = .005$  and  $m = 15$ . CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

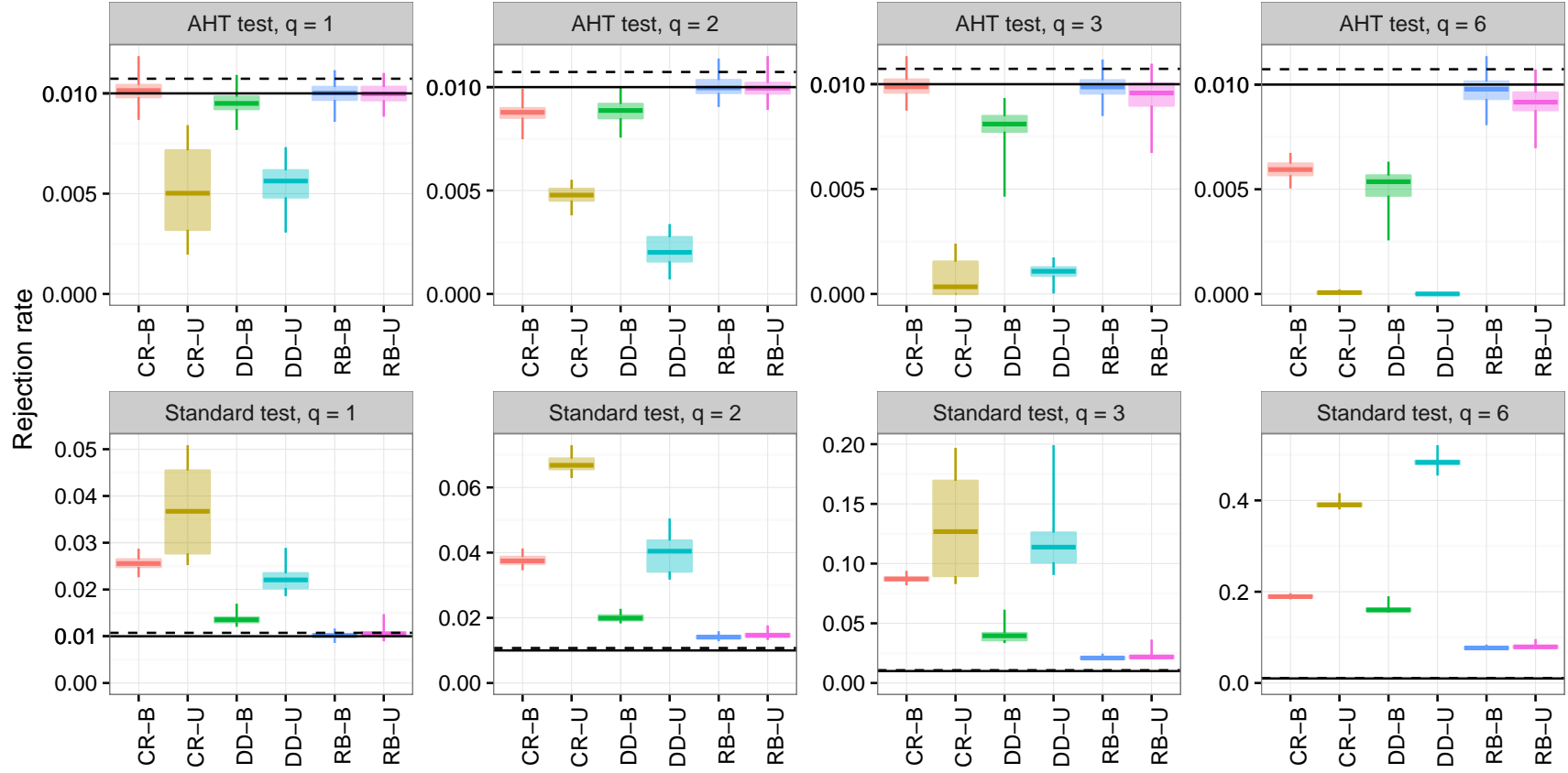


Figure S5: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis ( $q$ ) for  $\alpha = .01$  and  $m = 15$ . CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

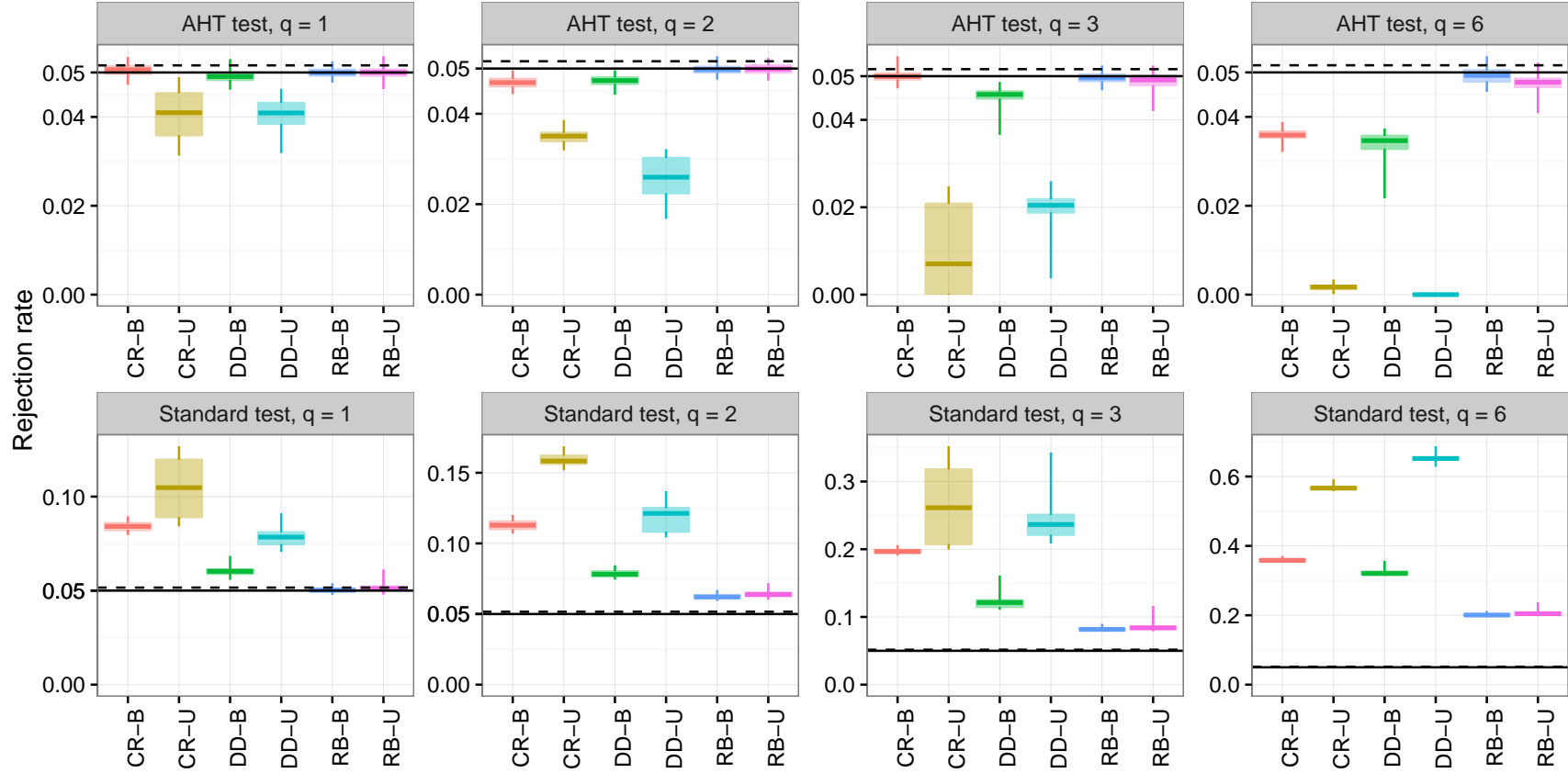


Figure S6: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis ( $q$ ) for  $\alpha = .05$  and  $m = 15$ . CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

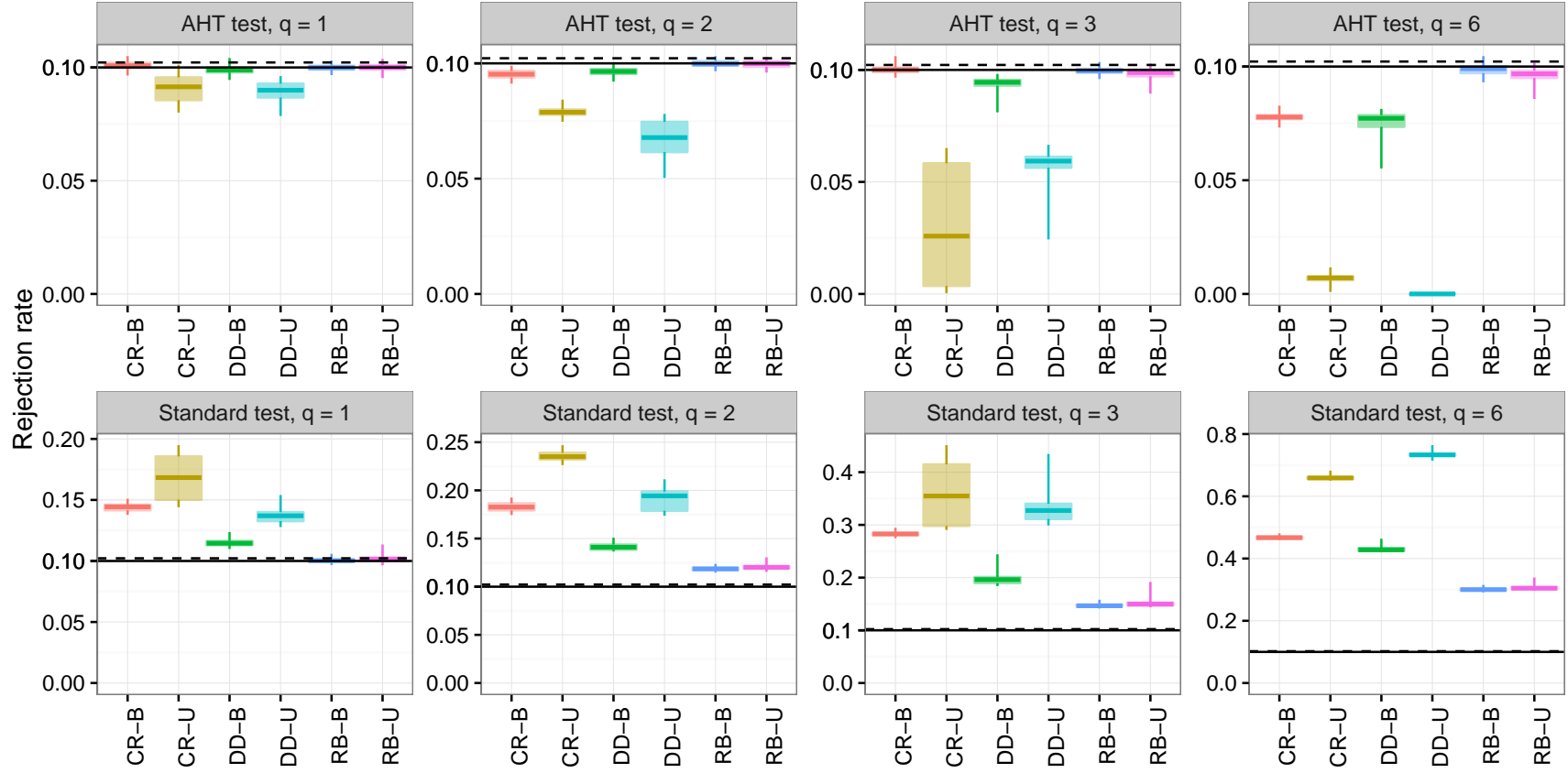


Figure S7: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis ( $q$ ) for  $\alpha = .10$  and  $m = 15$ . CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

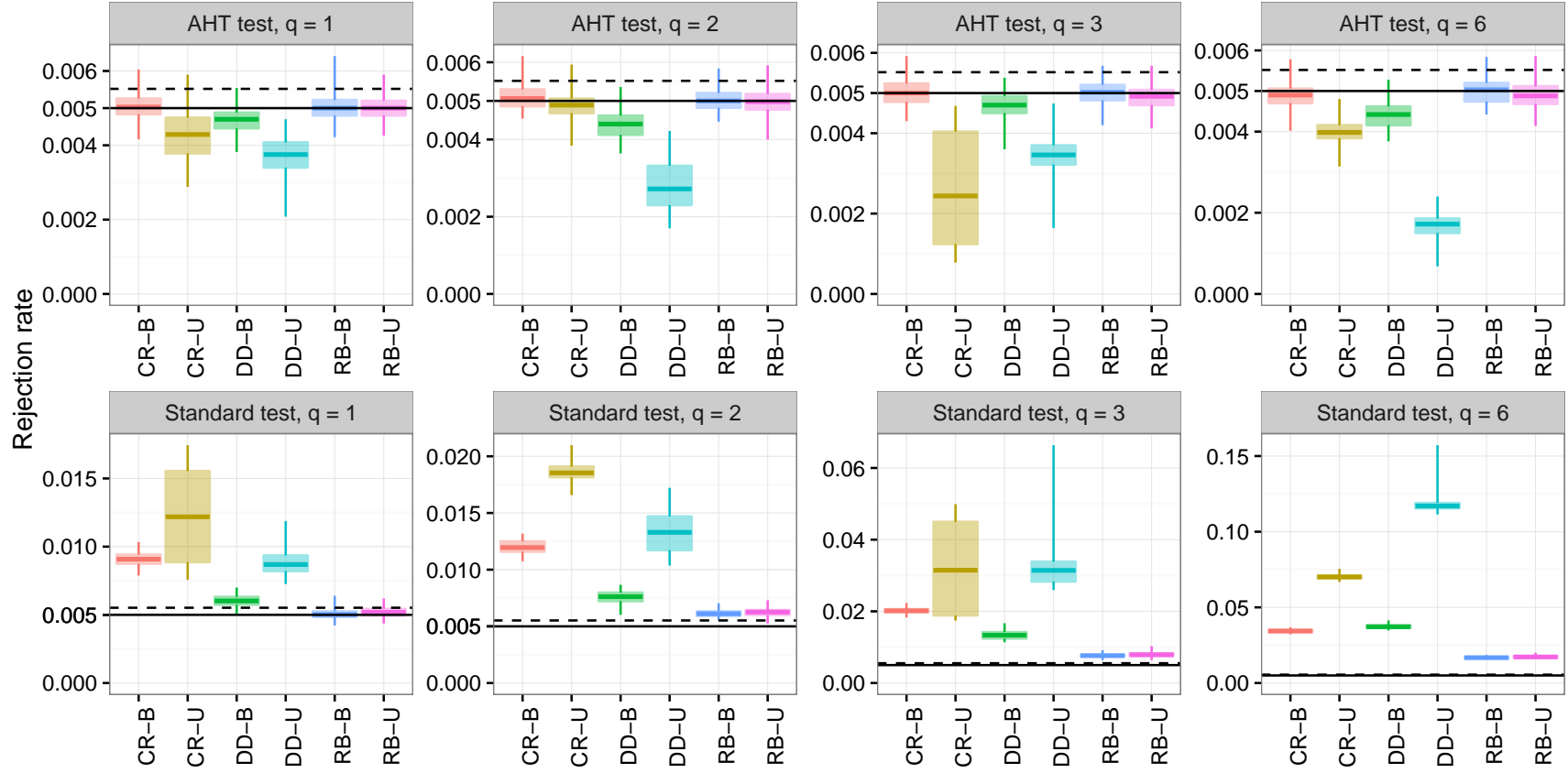


Figure S8: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis ( $q$ ) for  $\alpha = .005$  and  $m = 30$ . CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

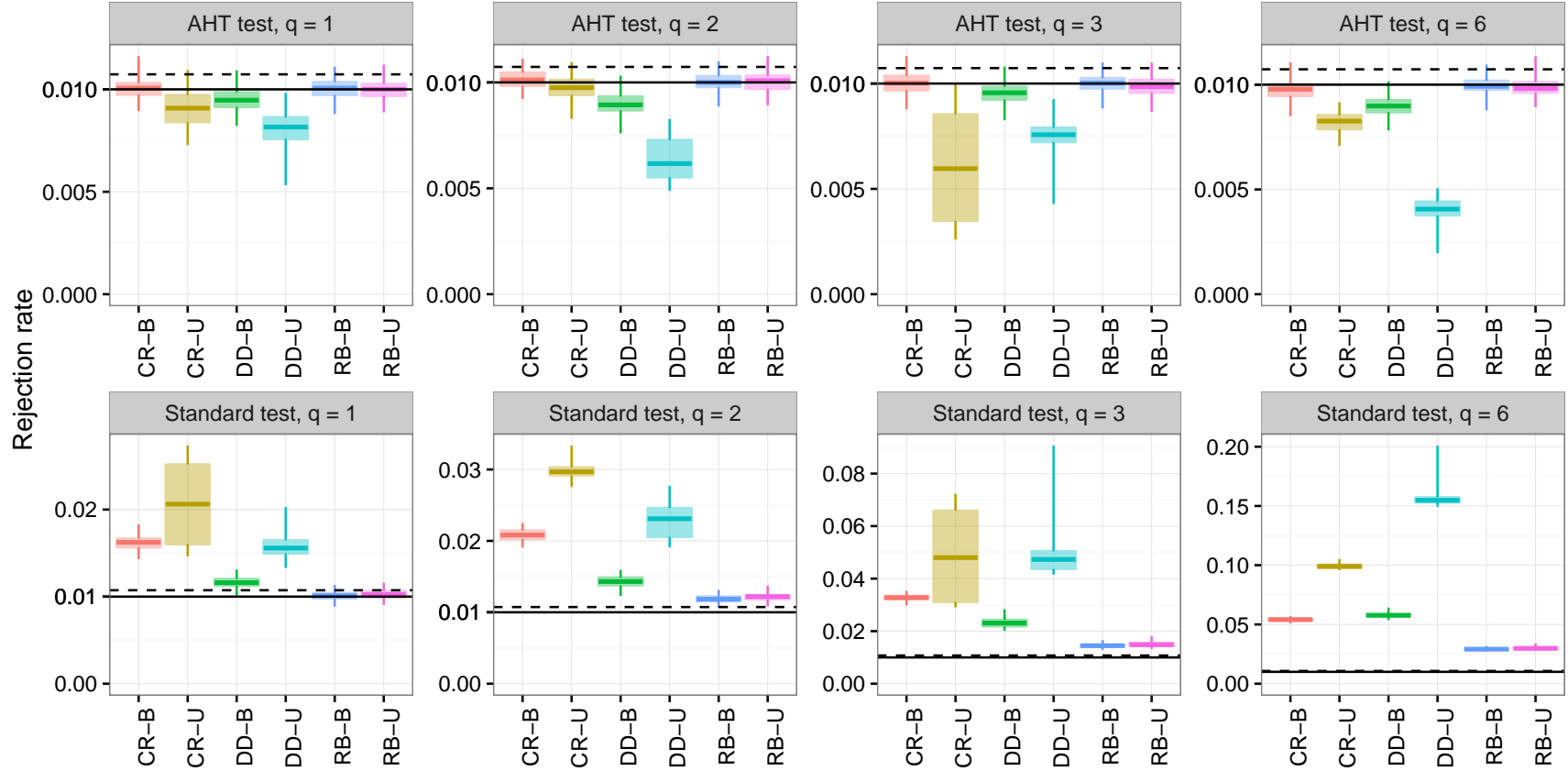


Figure S9: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis ( $q$ ) for  $\alpha = .01$  and  $m = 30$ . CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

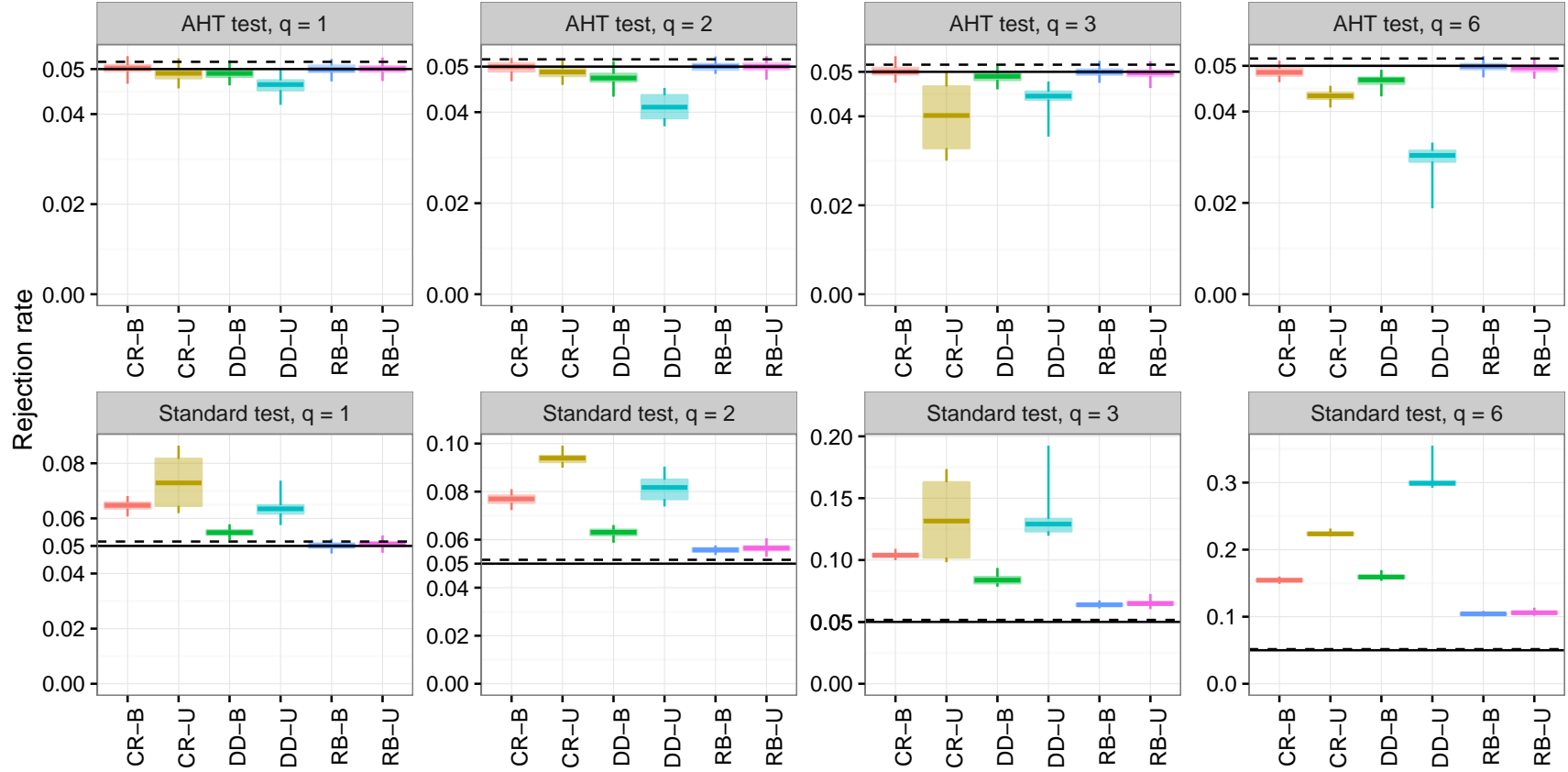


Figure S10: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis ( $q$ ) for  $\alpha = .05$  and  $m = 30$ . CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

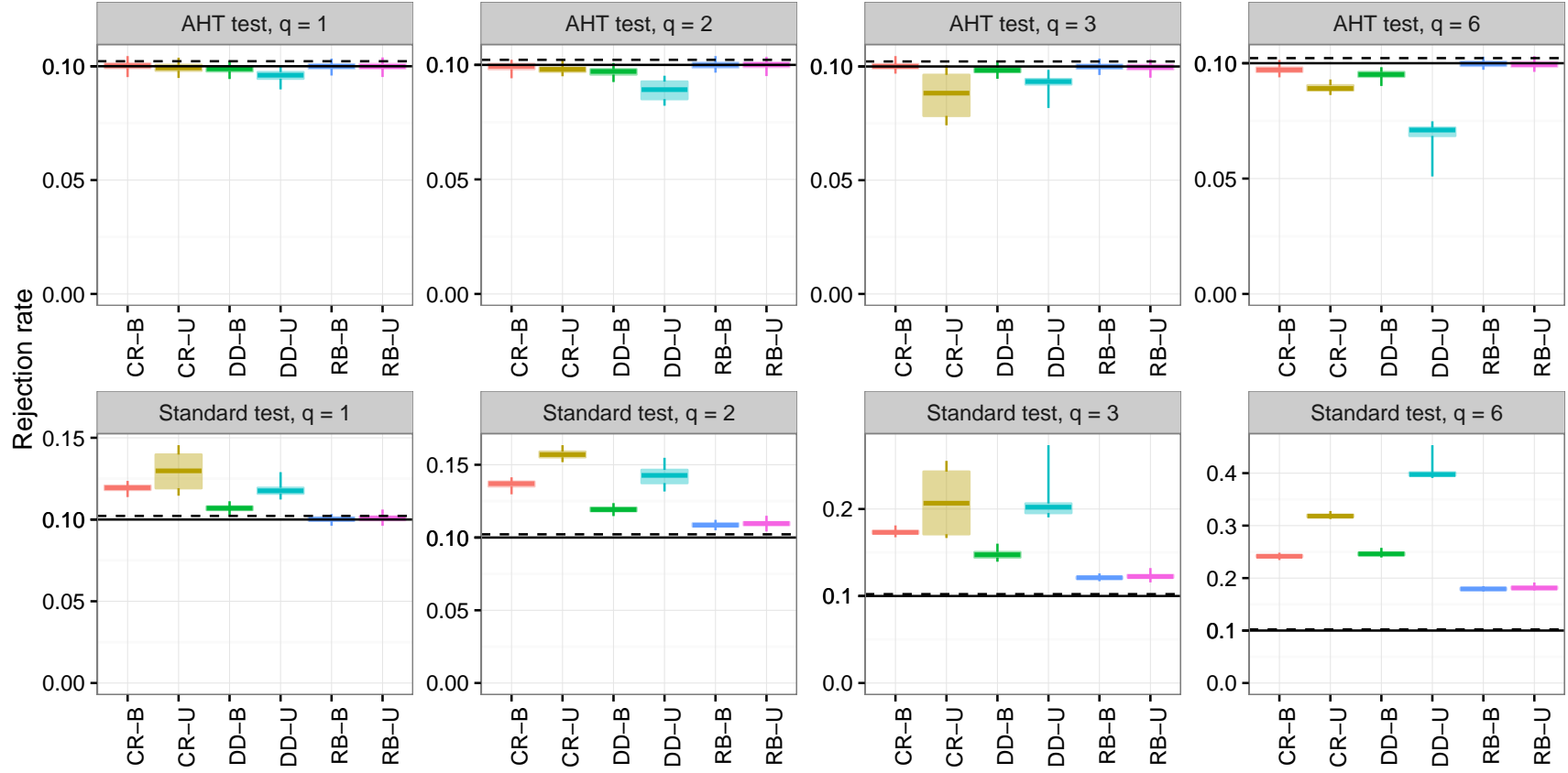


Figure S11: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis ( $q$ ) for  $\alpha = .10$  and  $m = 30$ . CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.



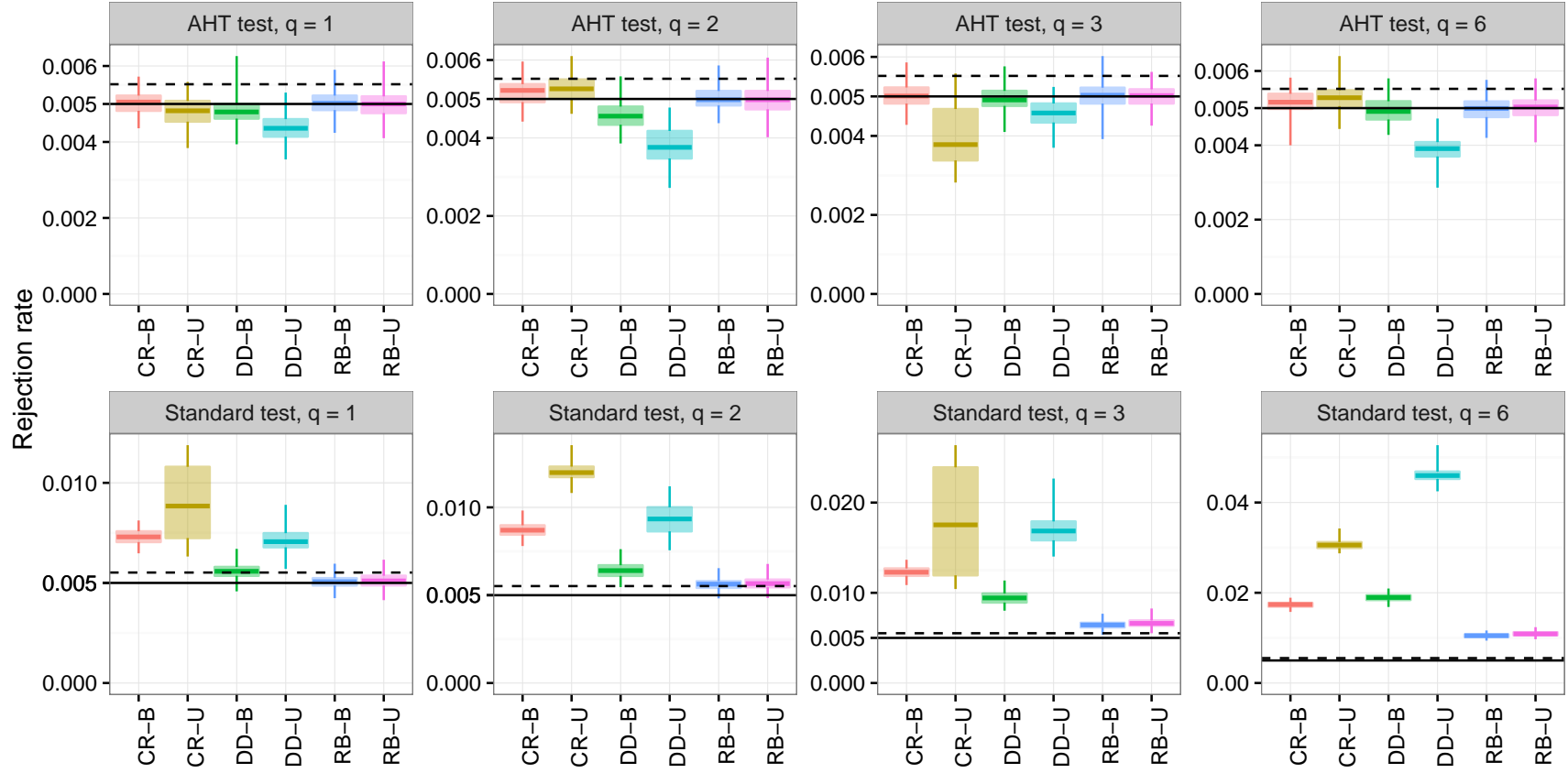


Figure S12: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis ( $q$ ) for  $\alpha = .005$  and  $m = 50$ . CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

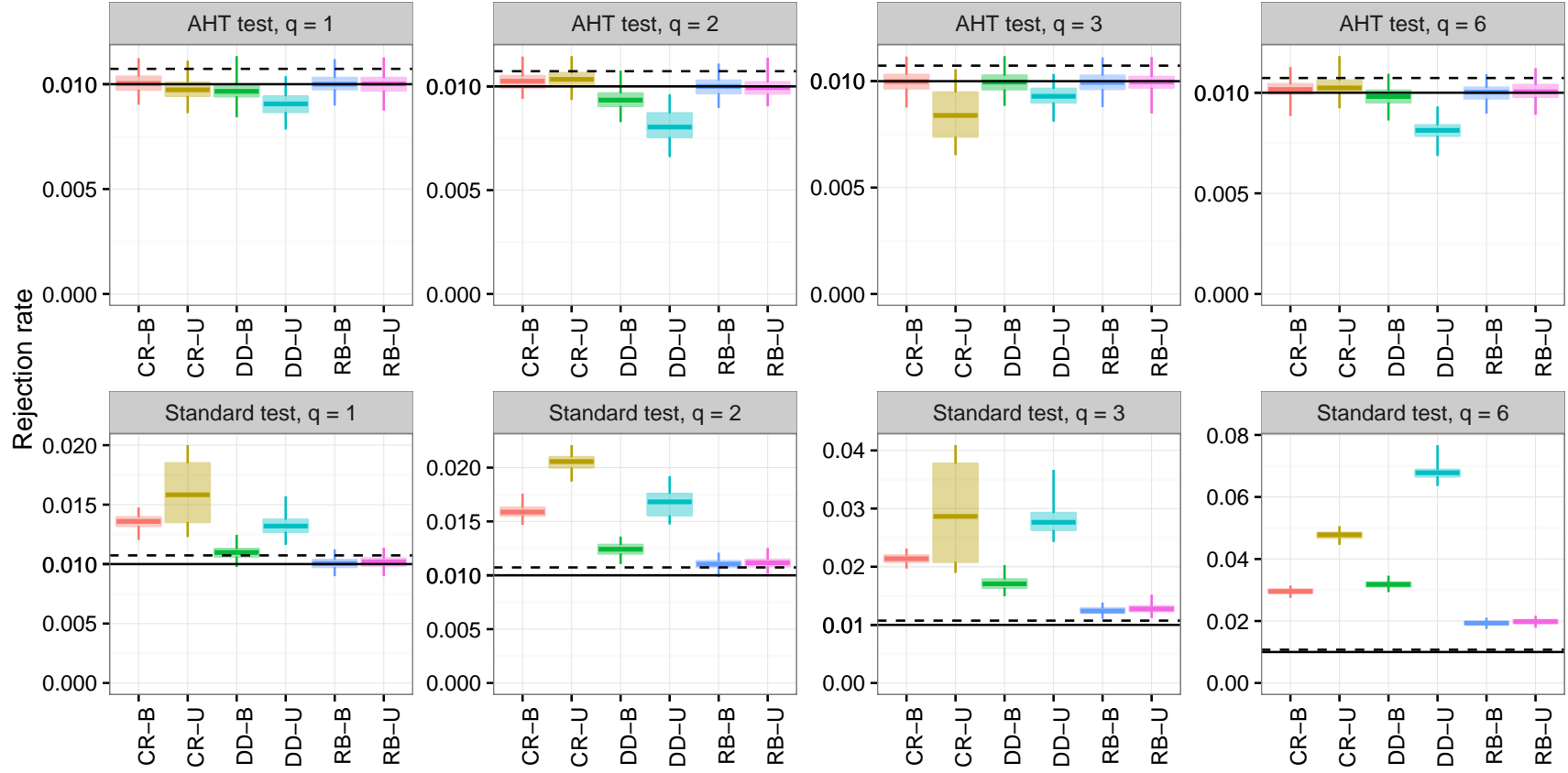


Figure S13: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis ( $q$ ) for  $\alpha = .01$  and  $m = 50$ . CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

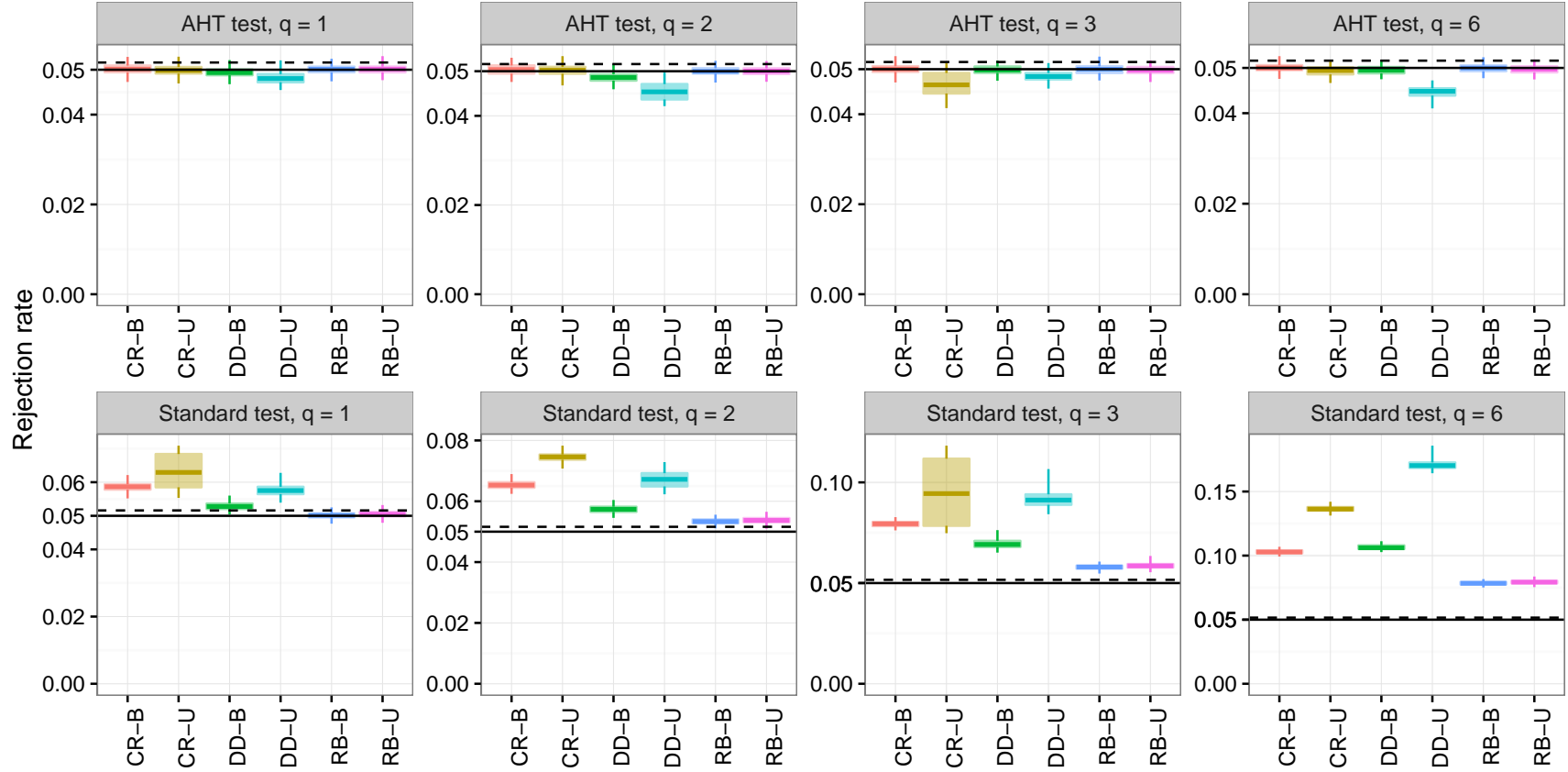


Figure S14: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis ( $q$ ) for  $\alpha = .05$  and  $m = 50$ . CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

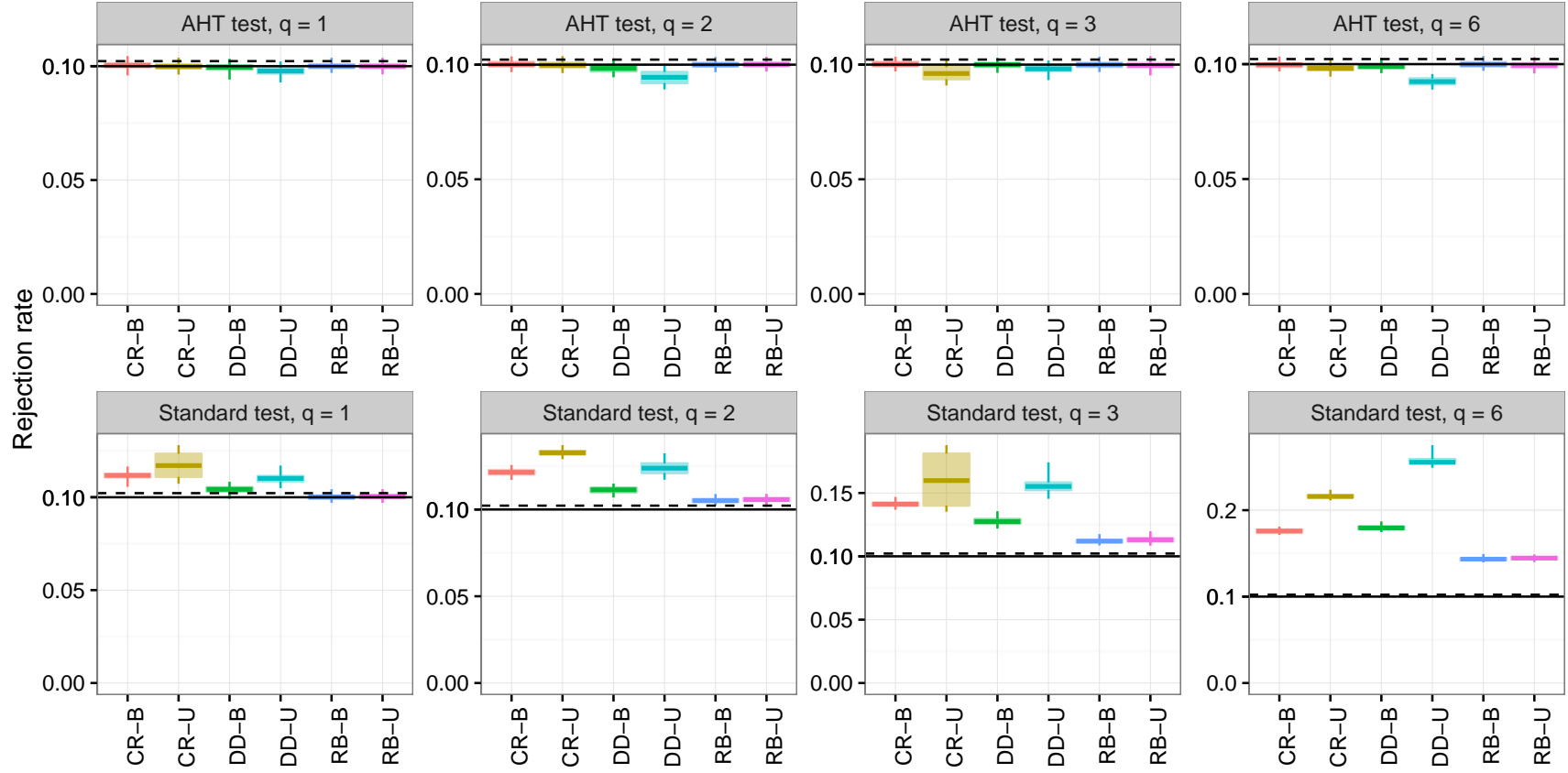


Figure S15: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis ( $q$ ) for  $\alpha = .10$  and  $m = 50$ . CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

### S4.3 Rejection rates of AHT test using CR1 or CR2, with and without accounting for absorption

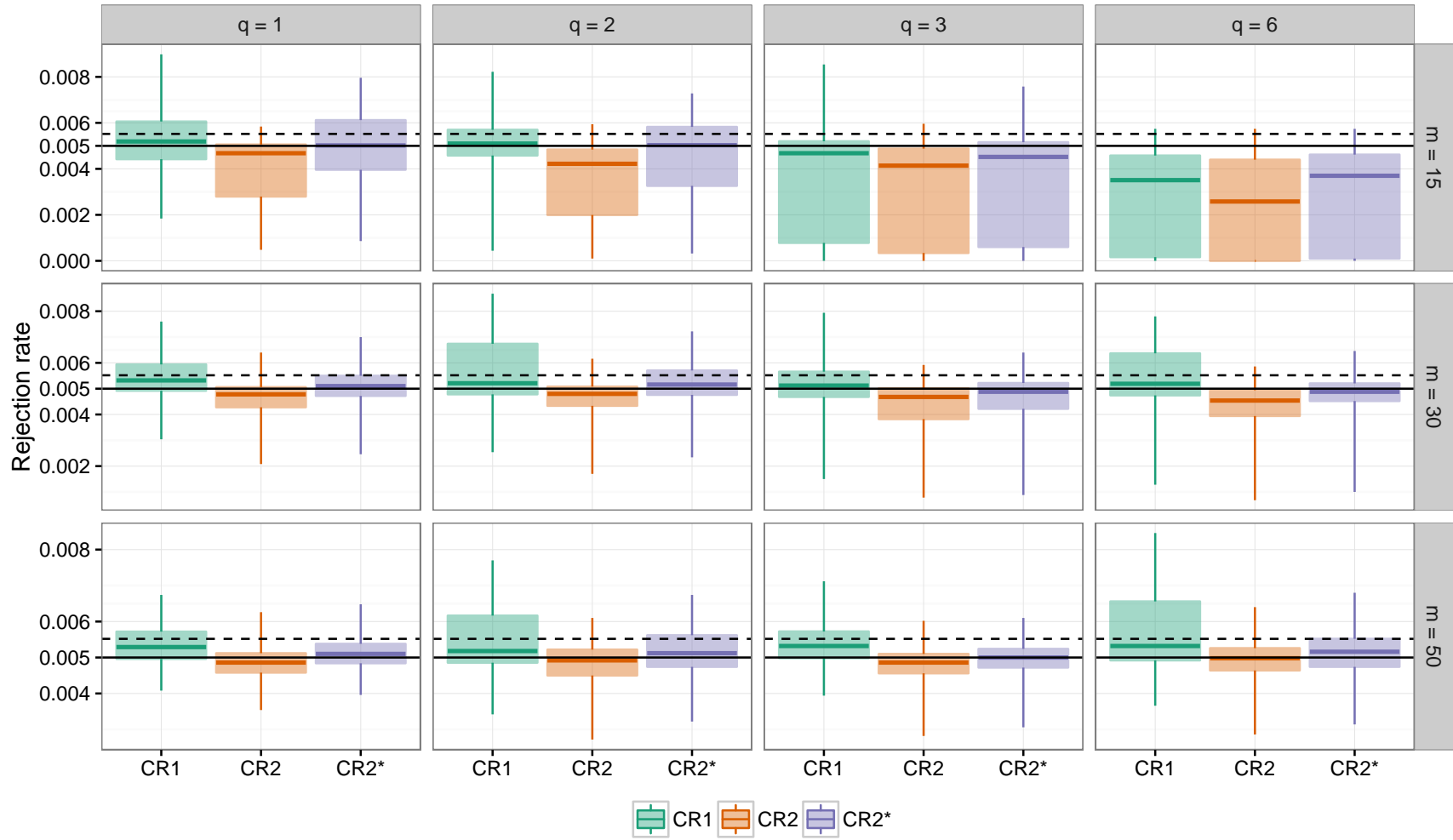


Figure S16: Rejection rates of AHT tests using CR1, CR2, or CR2 calculated without accounting for absorption of fixed effects (CR2\*), by sample size ( $m$ ) and dimension of hypothesis ( $q$ ), for  $\alpha = .005$ .

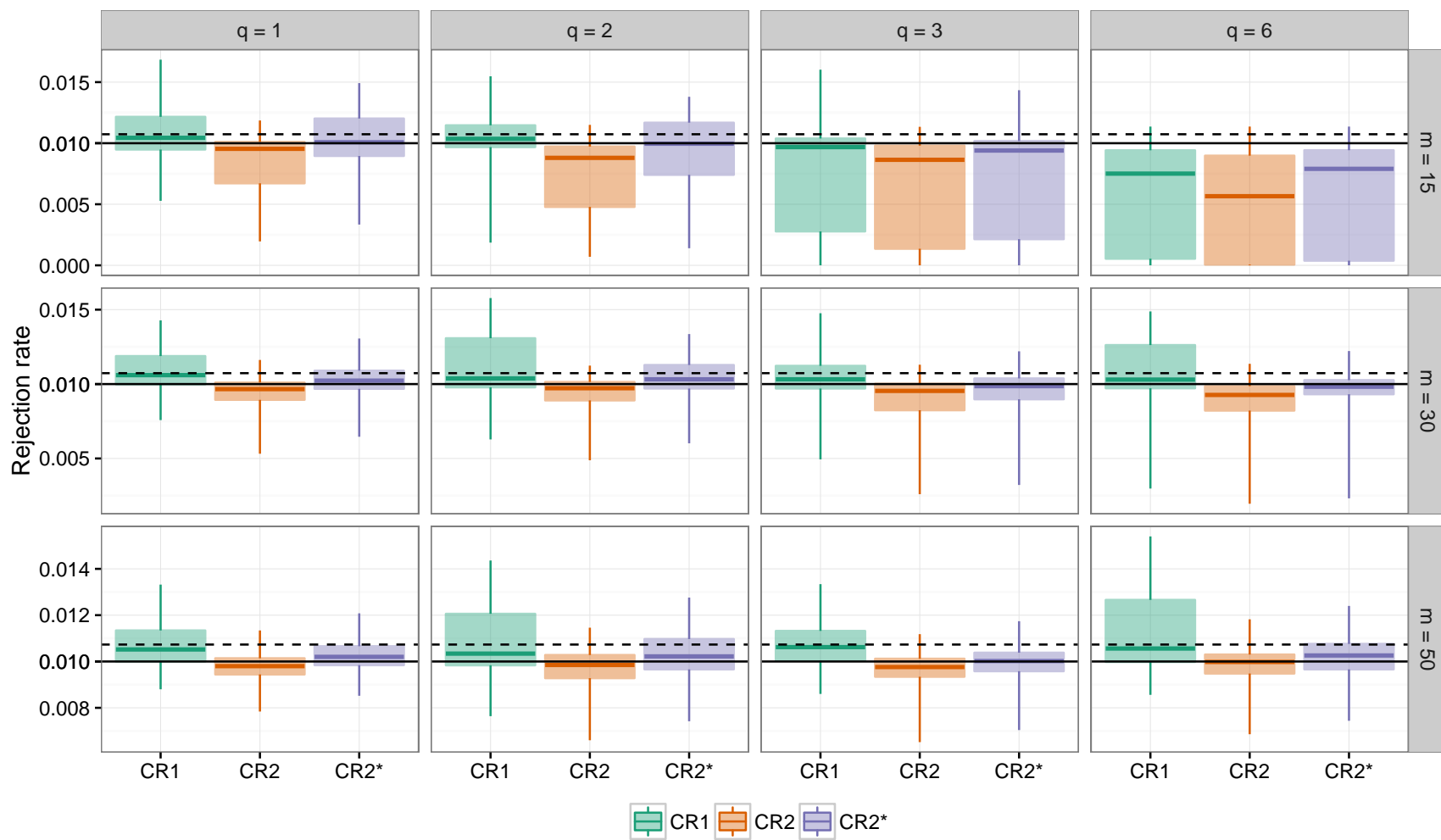


Figure S17: Rejection rates of AHT tests using CR1, CR2, or CR2 calculated without accounting for absorption of fixed effects (CR2\*), by sample size ( $m$ ) and dimension of hypothesis ( $q$ ), for  $\alpha = .01$ .

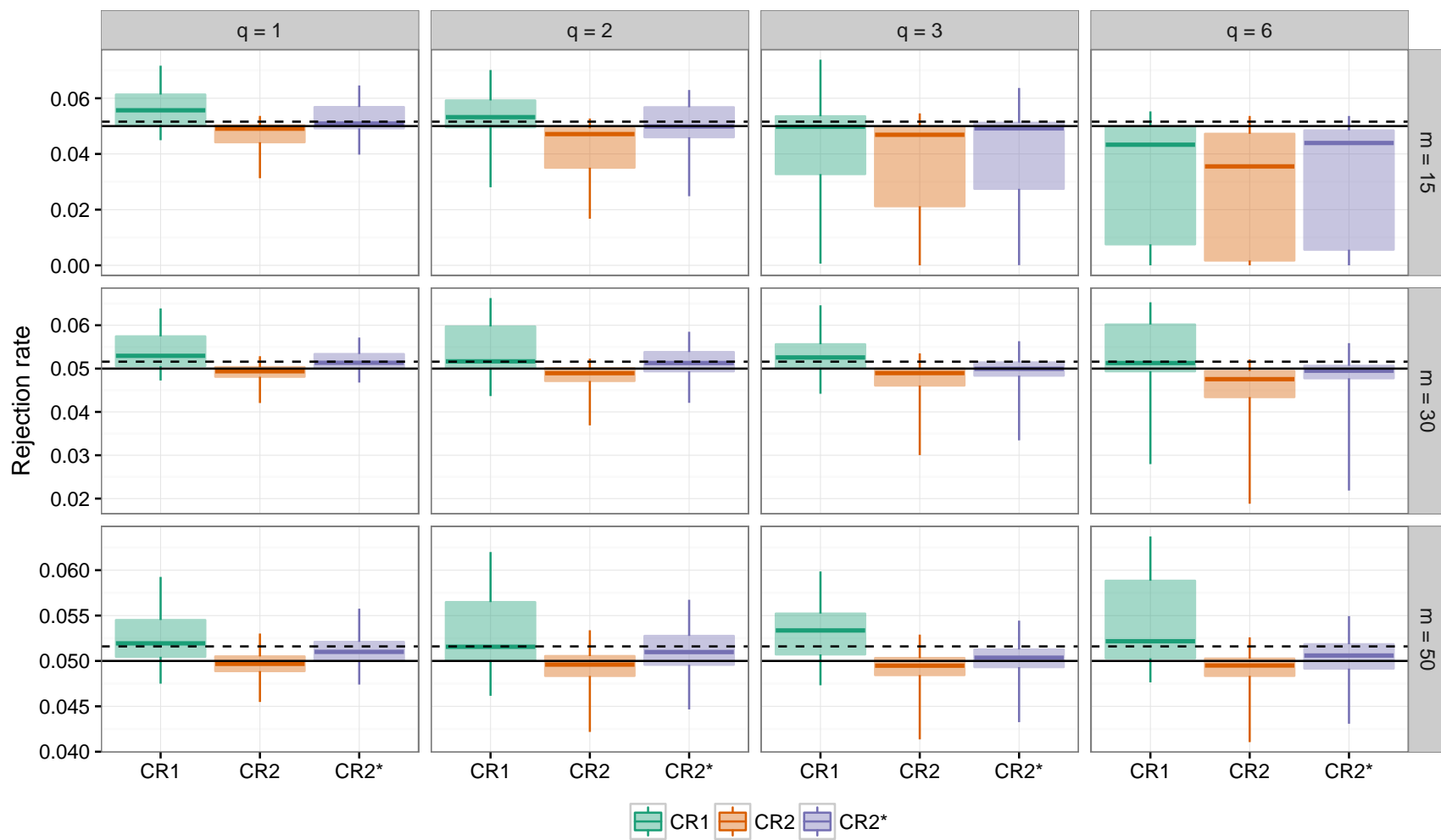


Figure S18: Rejection rates of AHT tests using CR1, CR2, or CR2 calculated without accounting for absorption of fixed effects (CR2\*), by sample size ( $m$ ) and dimension of hypothesis ( $q$ ), for  $\alpha = .05$ .

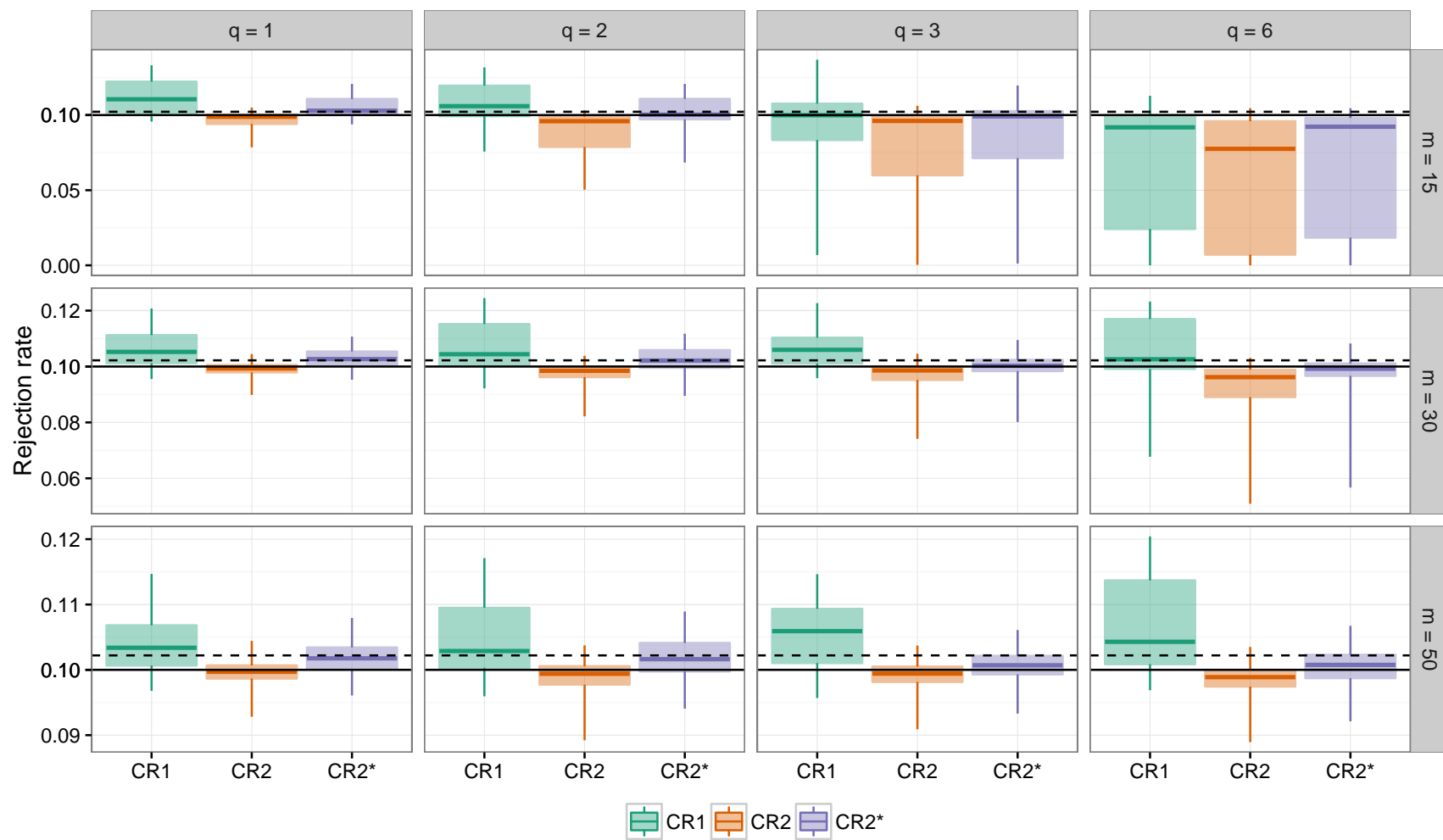


Figure S19: Rejection rates of AHT tests using CR1, CR2, or CR2 calculated without accounting for absorption of fixed effects (CR2\*), by sample size ( $m$ ) and dimension of hypothesis ( $q$ ), for  $\alpha = .10$ .



#### S4.4 Rejection rates of AHT test by degree of working model misspecification

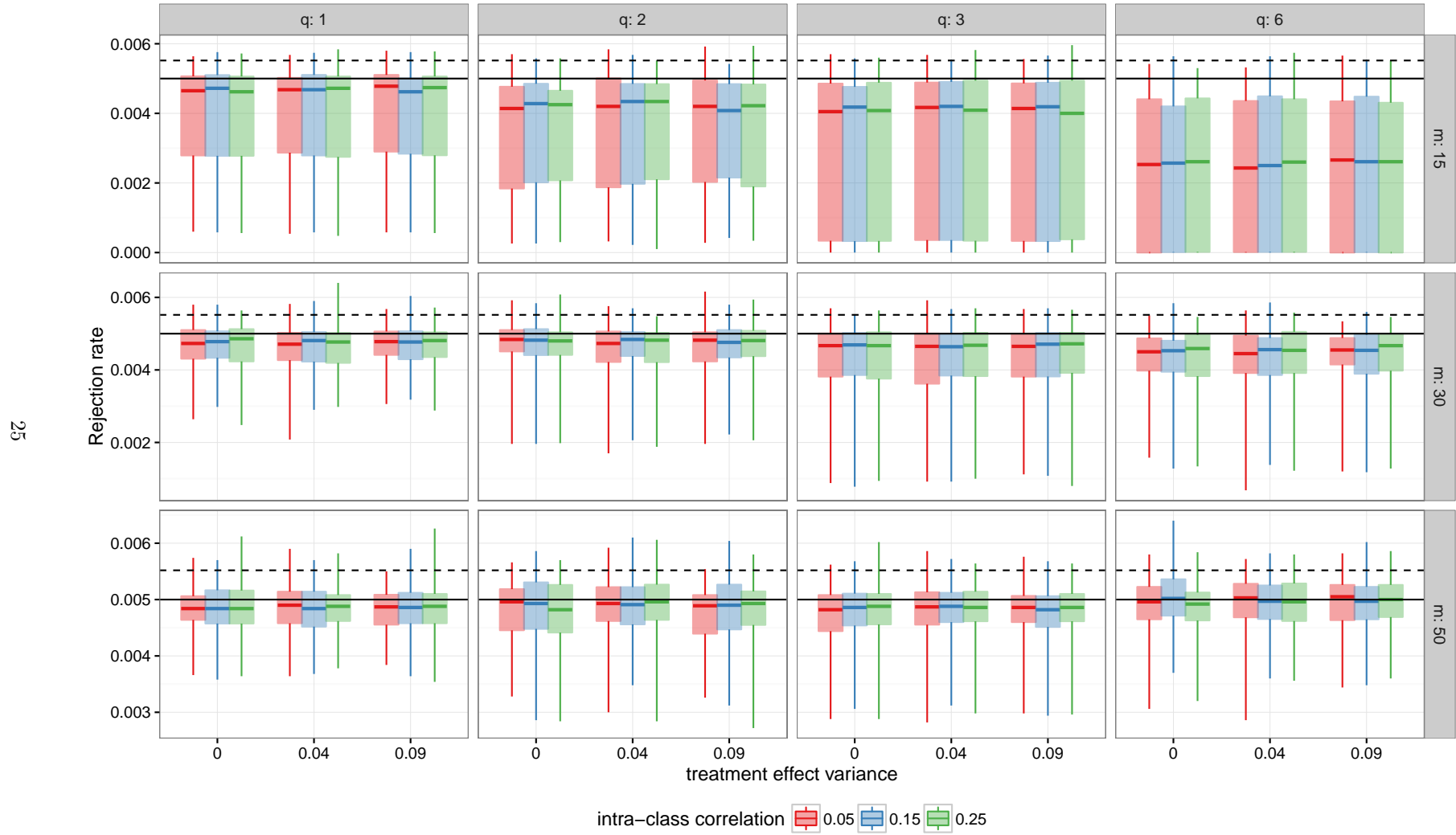


Figure S20: Rejection rates of CR2 AHT test, by treatment effect variance and intra-class correlation for  $\alpha = .005$ .

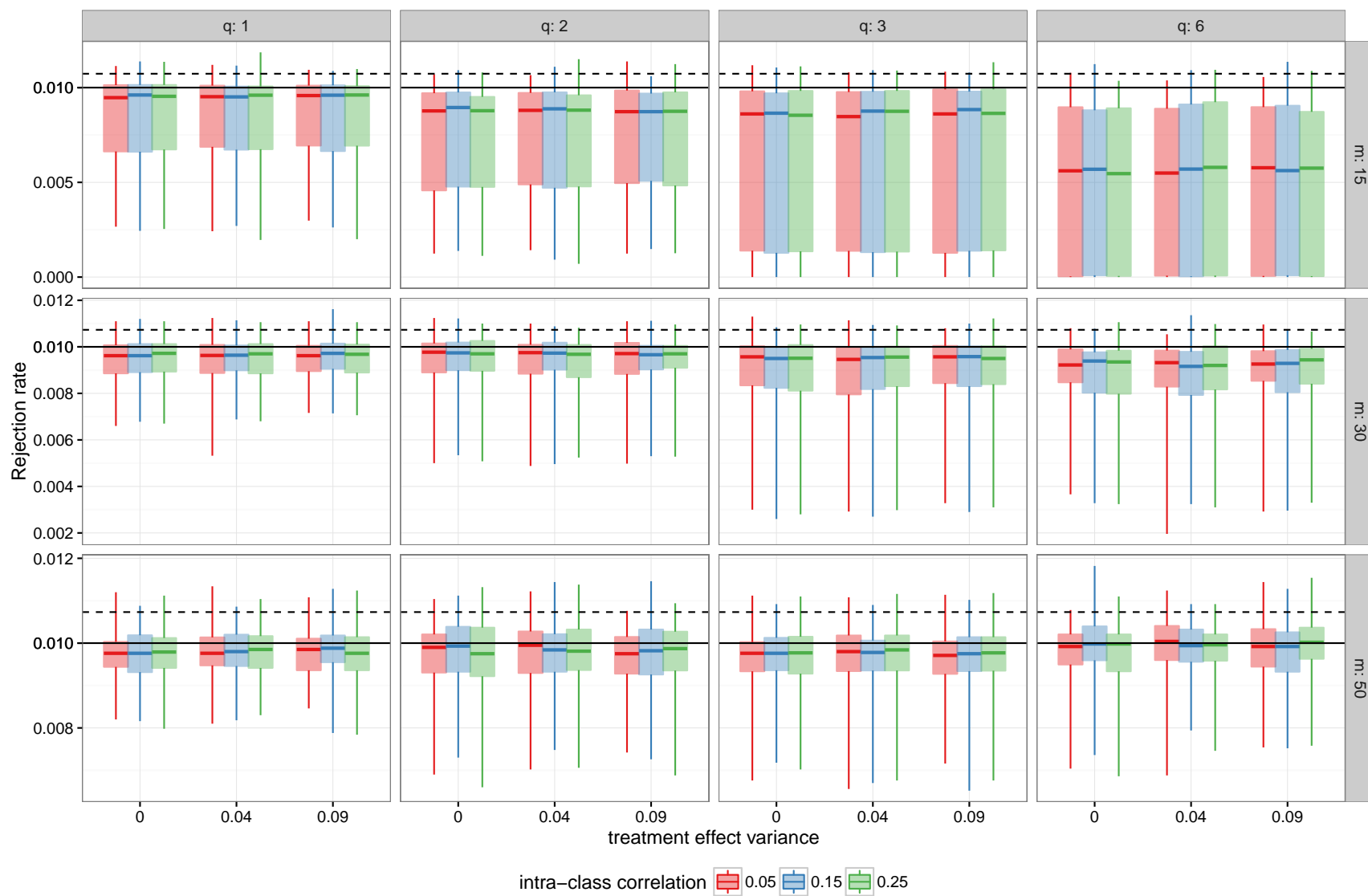


Figure S21: Rejection rates of CR2 AHT test, by treatment effect variance and intra-class correlation for  $\alpha = .01$ .

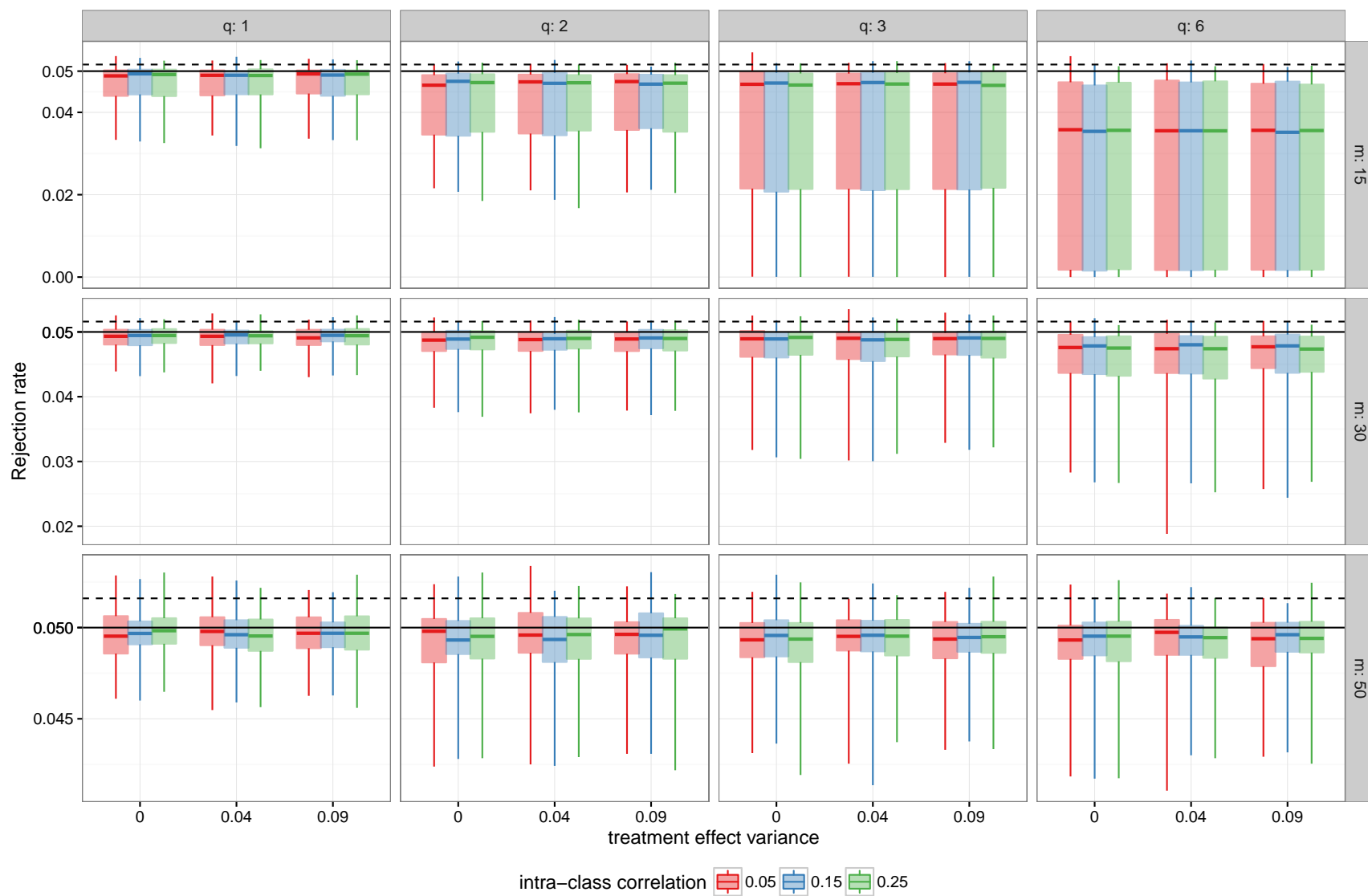


Figure S22: Rejection rates of CR2 AHT test, by treatment effect variance and intra-class correlation for  $\alpha = .05$ .

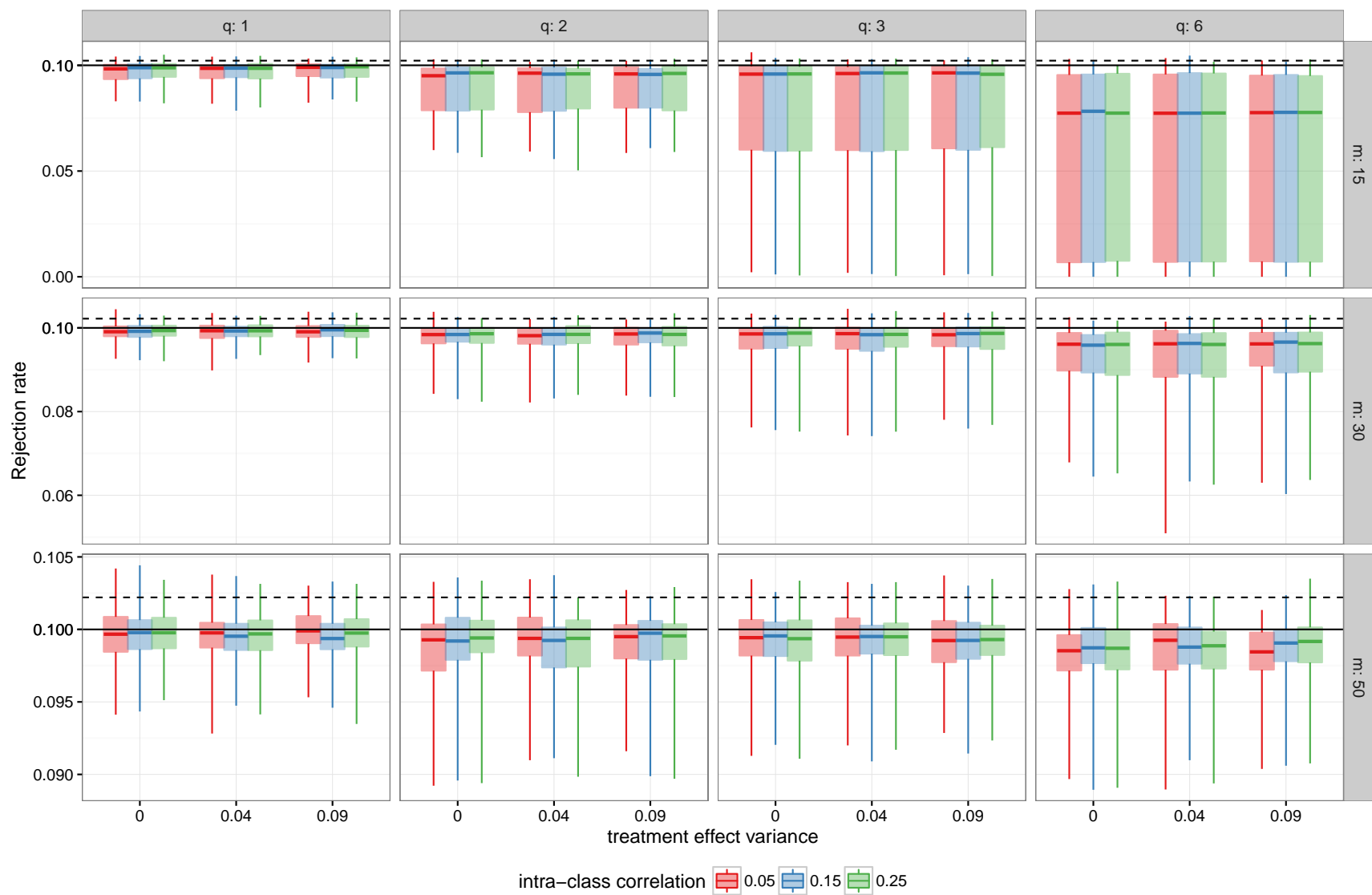


Figure S23: Rejection rates of CR2 AHT test, by treatment effect variance and intra-class correlation for  $\alpha = .10$ .

## References

Henderson, H. V. and Searle, S. R. (1981), ‘On deriving the inverse of a sum of matrices’, *Siam Review* **23**(1), 53–60.