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

July 1, 2016

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 $\mathbf{\Lambda}$ 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 $\mathbf{B}_i = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$. Then $\mathbf{B}_i^+ = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}'$ and $\mathbf{B}_i^{+1/2} = \mathbf{V}\mathbf{\Lambda}^{-1/2}\mathbf{V}'$. Now, observe that

$$\begin{aligned}
\ddot{\mathbf{R}}_i'\mathbf{W}_i\mathbf{A}_i\left(\mathbf{I}-\mathbf{H_X}\right)_i\mathbf{\Phi}\left(\mathbf{I}-\mathbf{H_X}\right)_i'\mathbf{A}_i'\mathbf{W}_i\ddot{\mathbf{R}}_i &= \ddot{\mathbf{R}}_i'\mathbf{W}_i\mathbf{D}_i\mathbf{B}_i^{+1/2}\mathbf{B}_i\mathbf{B}_i^{+1/2}\mathbf{D}_i'\mathbf{W}_i\ddot{\mathbf{R}}_i \\
&= \ddot{\mathbf{R}}_i'\mathbf{W}_i\mathbf{D}_i\mathbf{V}\mathbf{V}'\mathbf{D}_i'\mathbf{W}_i\ddot{\mathbf{R}}_i.
\end{aligned} \tag{1}$$

Because $\mathbf{D}_i$, and $\mathbf{\Phi}$ are positive definite and $\mathbf{B}_i$ is symmetric, the eigen-vectors $\mathbf{V}$ define an orthonormal basis for the column span of $\left(\mathbf{I}-\mathbf{H_X}\right)_i$. We now show that $\ddot{\mathbf{U}}_i$ is in the column space of $\left(\mathbf{I}-\mathbf{H_X}\right)_i$. 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 $\mathbf{Z} = \left(\mathbf{Z}_1', ..., \mathbf{Z}_m'\right)'$. Now observe that $\left(\mathbf{I}-\mathbf{H_T}\right)\mathbf{Z} = \mathbf{Z}$. It follows that

$$\begin{aligned}
\left(\mathbf{I}-\mathbf{H_X}\right)_i\mathbf{Z} = \left(\mathbf{I}-\mathbf{H}_{\ddot{\mathbf{U}}}\right)_i\left(\mathbf{I}-\mathbf{H_T}\right)\mathbf{Z} &= \left(\mathbf{I}-\mathbf{H}_{\ddot{\mathbf{U}}}\right)_i\mathbf{Z} \\
&= \mathbf{Z}_i - \ddot{\mathbf{U}}_i\mathbf{M}_{\ddot{\mathbf{U}}}\sum_{k=1}^m \ddot{\mathbf{U}}_k'\mathbf{W}_k\mathbf{Z}_k \\
&= \ddot{\mathbf{U}}_i\mathbf{M}_{\ddot{\mathbf{U}}}\left(\sum_{k\neq i}\ddot{\mathbf{U}}_k'\mathbf{W}_k\ddot{\mathbf{U}}\right)\mathbf{L}^{-1}\mathbf{M}_{\ddot{\mathbf{U}}}^{-1} = \ddot{\mathbf{U}}_i.
\end{aligned}$$

Thus, there exists an $N \times (r+s)$ matrix $\mathbf{Z}$ such that $\left(\mathbf{I}-\mathbf{H}_{\ddot{\mathbf{X}}}\right)_i\mathbf{Z} = \ddot{\mathbf{U}}_i$, i.e., $\ddot{\mathbf{U}}_i$ is in the column span of $\left(\mathbf{I}-\mathbf{H_X}\right)_i$. 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 $\left(\mathbf{I}-\mathbf{H_X}\right)_i$. It follows that

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

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

$$\text{Var}\left(\ddot{\mathbf{e}}_i\right) = \left(\mathbf{I}-\mathbf{H_X}\right)_i\mathbf{\Phi}\left(\mathbf{I}-\mathbf{H_X}\right)_i',$$

It follows that

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

S2 Proof of Theorem 2

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

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

and

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

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

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

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

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

The last equality follows from the fact that

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

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

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

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

S3 Details of simulation study

This section provides further details regarding the design of the simulations reported in Section 4 of the main text. 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 pattern of treatment assignments. Let n_{ghi} denote the number of units at which cluster i in group g is observed under condition h, for $i = 1, ..., m$, $g = 1, ..., G$, and $h = 1, 2, 3$. The following six designs were simulated:

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, with $G = 1$, $m_1 = m$, $n_{11i} = n/2$, $n_{12i} = n/3$, and $n_{13i} = n/6$.

2. An unbalanced, block-randomized design, with two different patterns of treatment allocation. Here, $G = 2$, $m_1 = m_2 = m/2$, $n_{11i} = n/2$, $n_{12i} = n/3$, $n_{13i} = n/6$, $n_{21i} = n/3$, $n_{22i} = 5n/9$, and $n_{23i} = n/9$.

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. Here, $G = 3$, $m_g = m/3$, and $n_{ghi} = n$ for $g = h$ and zero otherwise.

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. Here, $G = 3$; $m_1 = 0.5m, m_2 = 0.3m, m_3 = 0.2m$; and $n_{ghi} = n$ for $g = h$ and zero otherwise.

5. A balanced difference-in-differences design, with two patterns of treatment allocation ($G = 2$) and clusters allocated equally to each pattern ($m_1 = m_2 = m/2$). Here, half of the clusters are observed under the first treatment condition only ($n_{11i} = n$) and the remaining half are observed under all three conditions, with $n_{21i} = n/2$, $n_{22i} = n/3$, and $n_{23i} = n/6$.

6. An unbalanced difference-in-differences design, again with two patterns of treatment allocation ($G = 2$), but where $m_1 = 2m/3$ clusters are observed under the first treatment condition only ($n_{11i} = n$) and the remaining $m_2 = m/3$ clusters are observed under all three conditions, with $n_{21i} = n/2$, $n_{22i} = n/3$, and $n_{23i} = n/6$.

S4 Additional simulation results

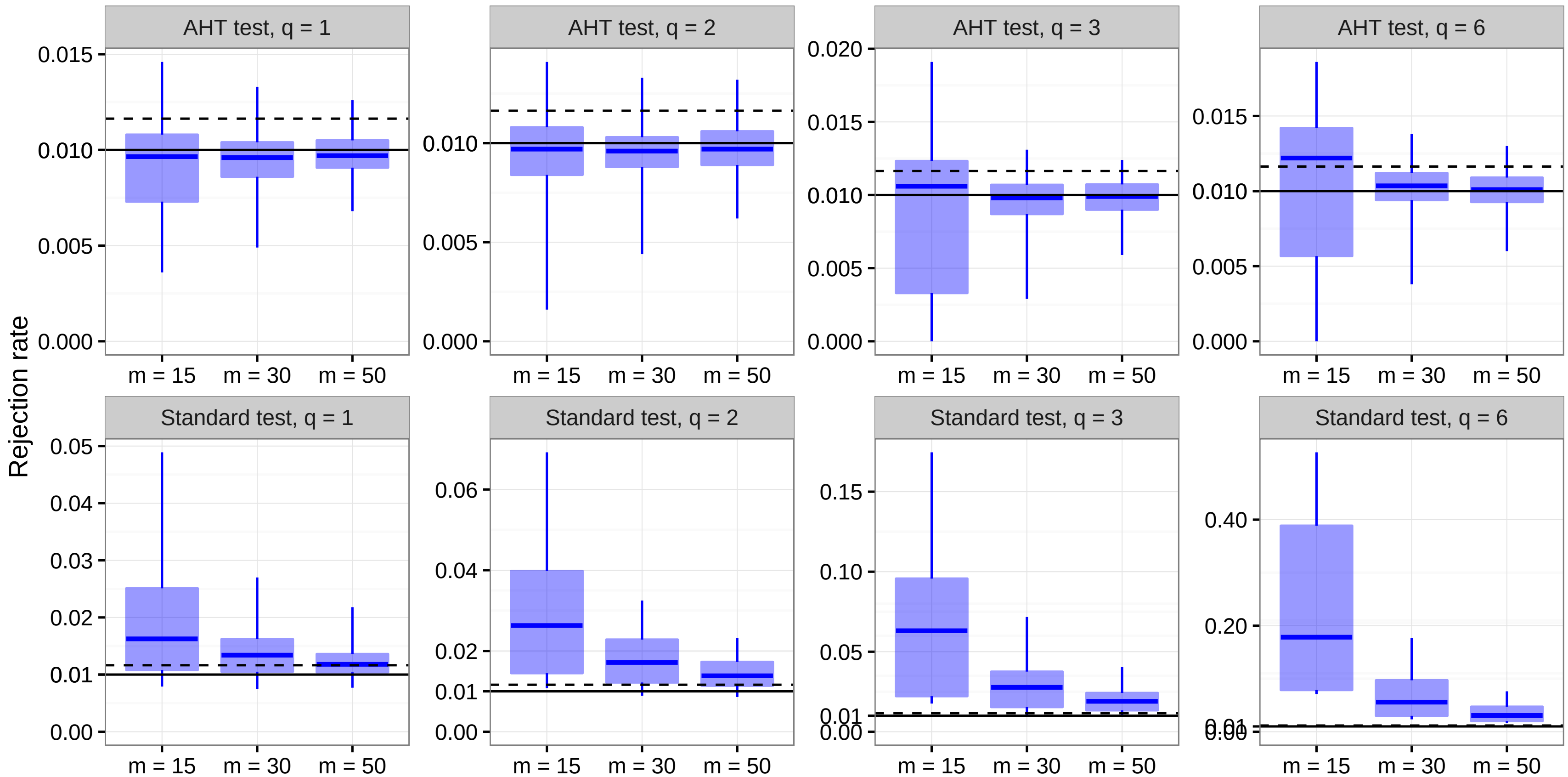

Figure S1: Rejection rates of AHT and standard tests for $\alpha = .01$, by dimension of hypothesis (q) and sample size (m).

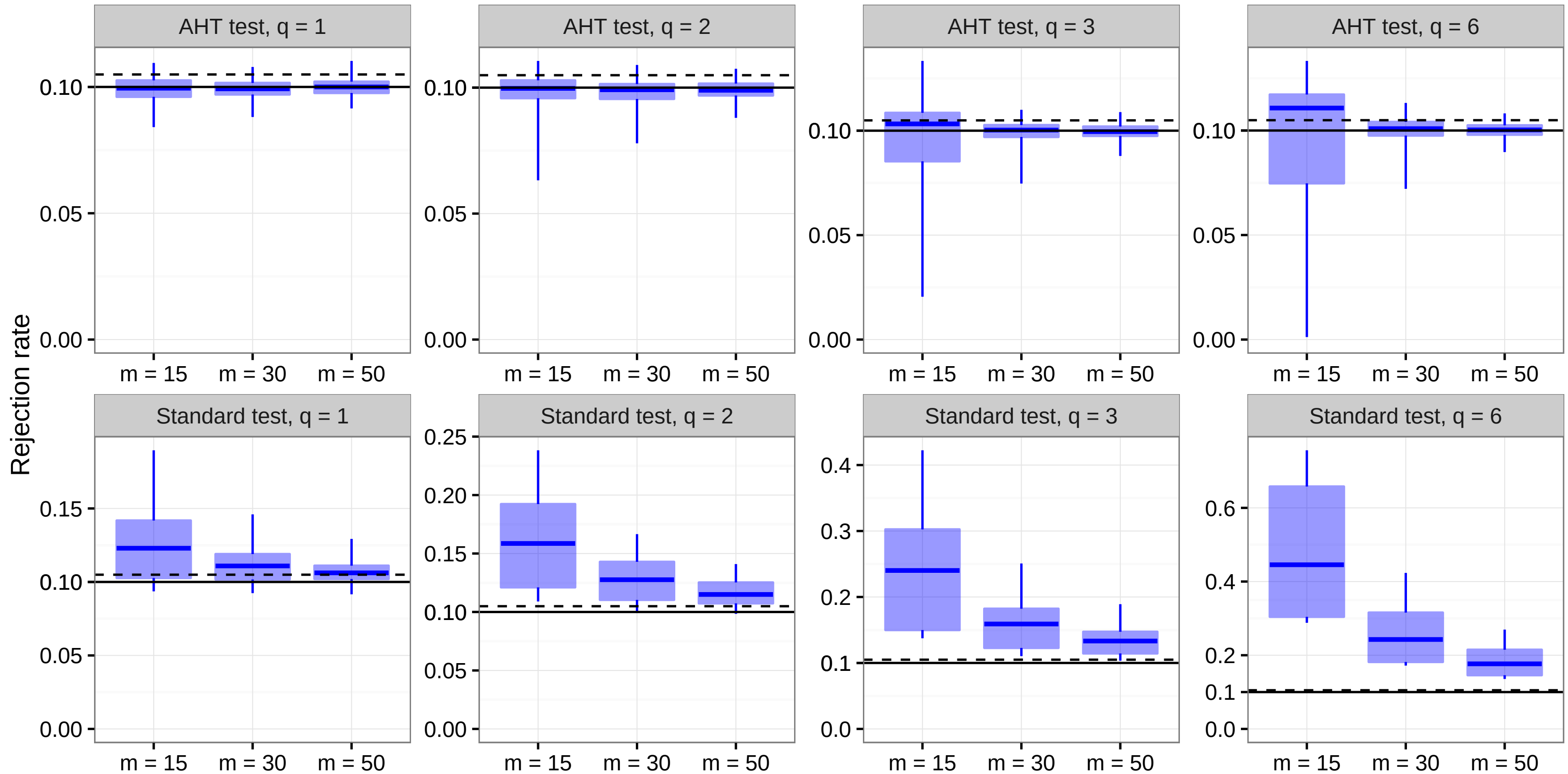

Figure S2: Rejection rates of AHT and standard tests for $\alpha = .10$, by dimension of hypothesis (q) and sample size (m).

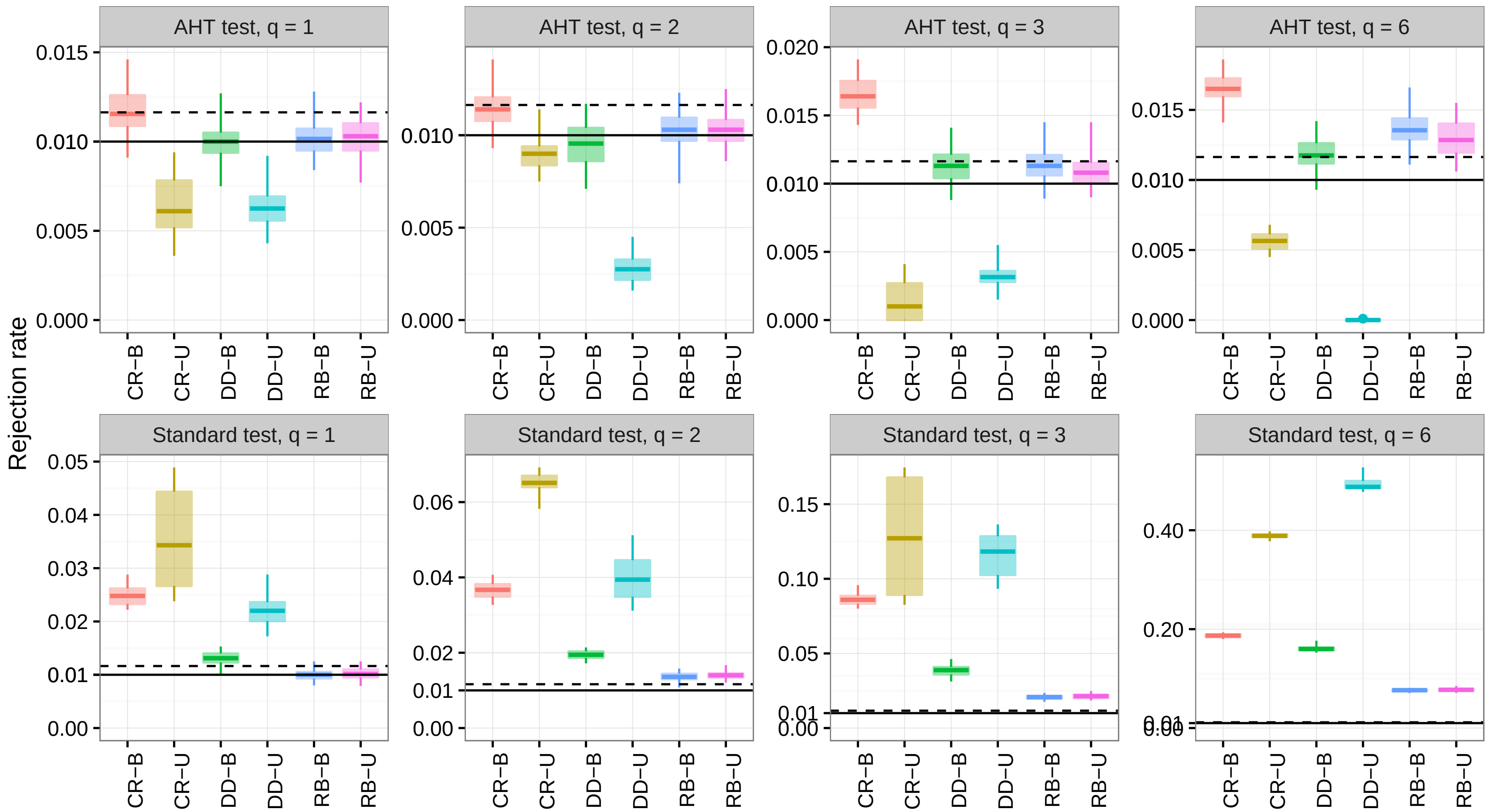

Figure S3: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis (q) for $\alpha = .01$ and $m = 15$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

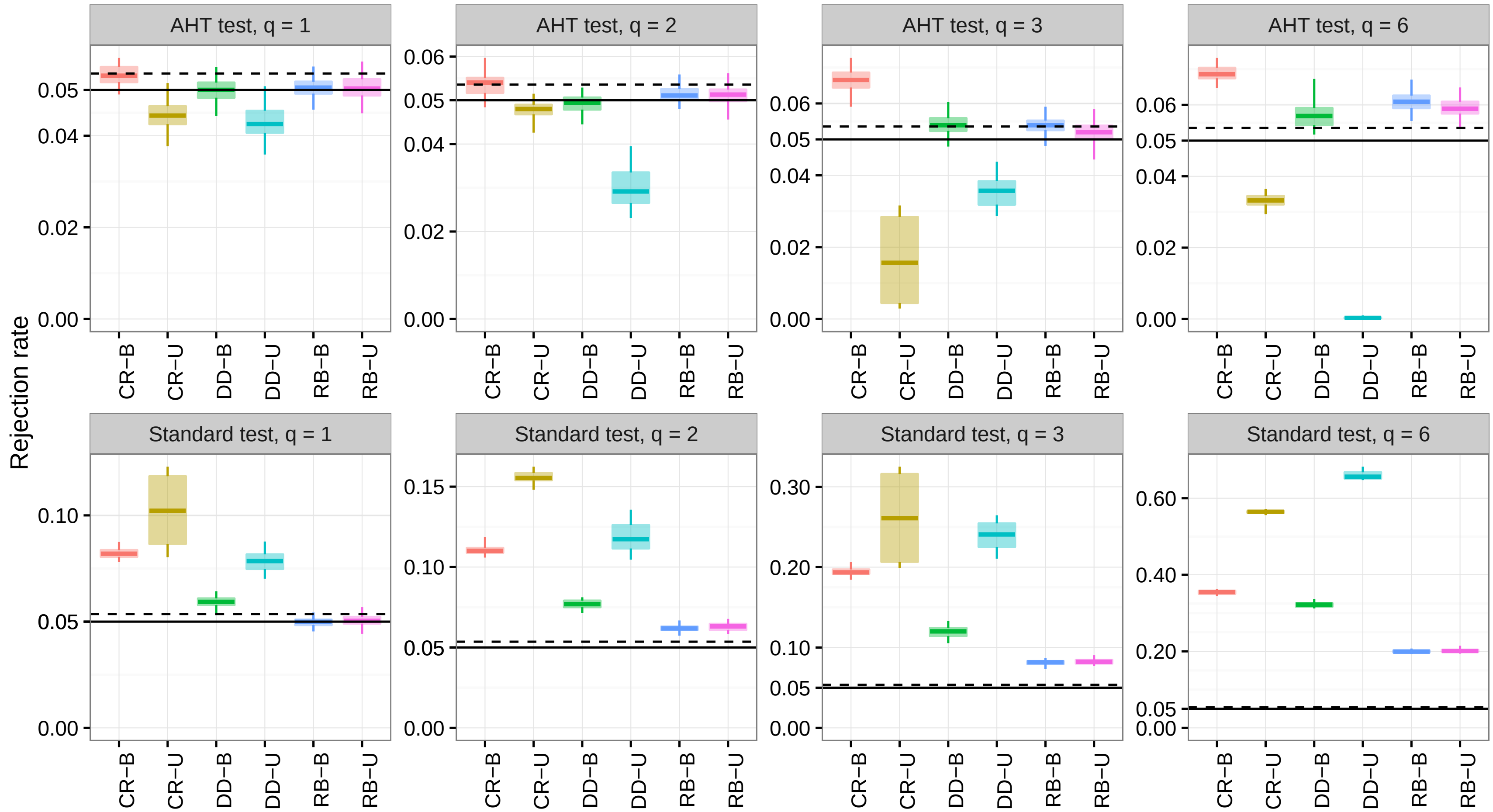

Figure S4: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis (q) for $\alpha = .05$ and $m = 15$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

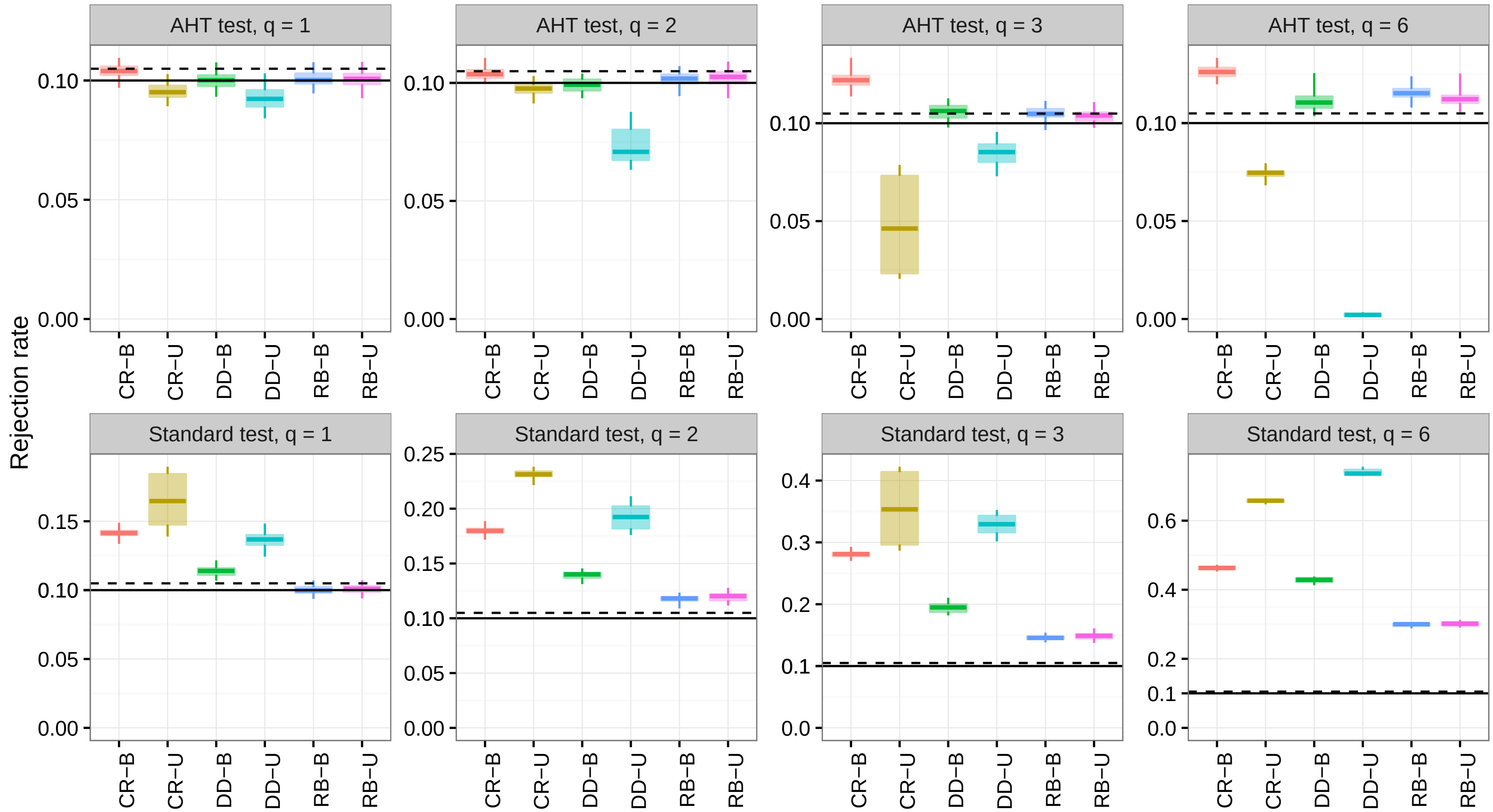

Figure S5: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis (q) for $\alpha = .10$ and $m = 15$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

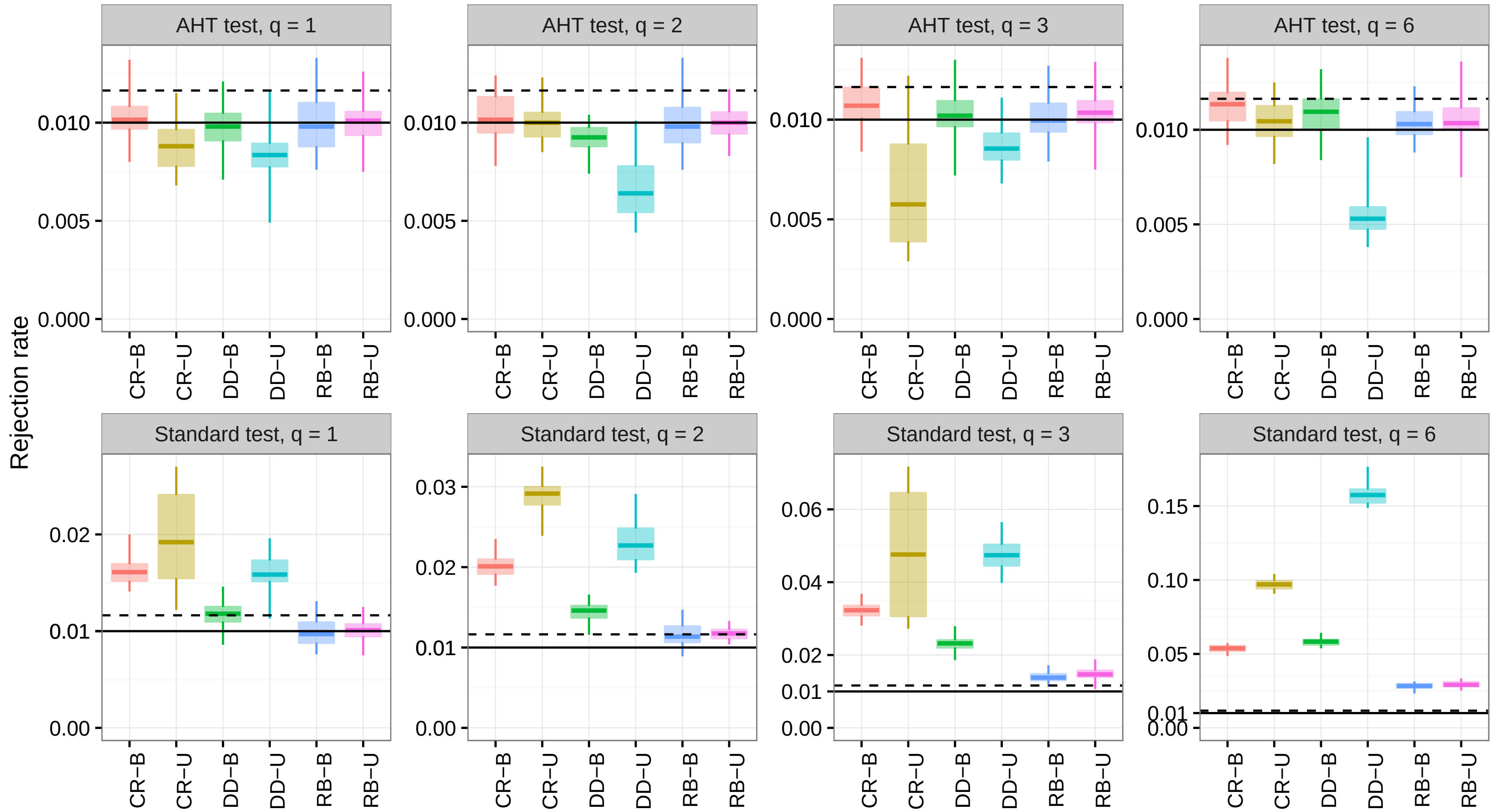

Figure S6: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis (q) for $\alpha = .01$ and $m = 30$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

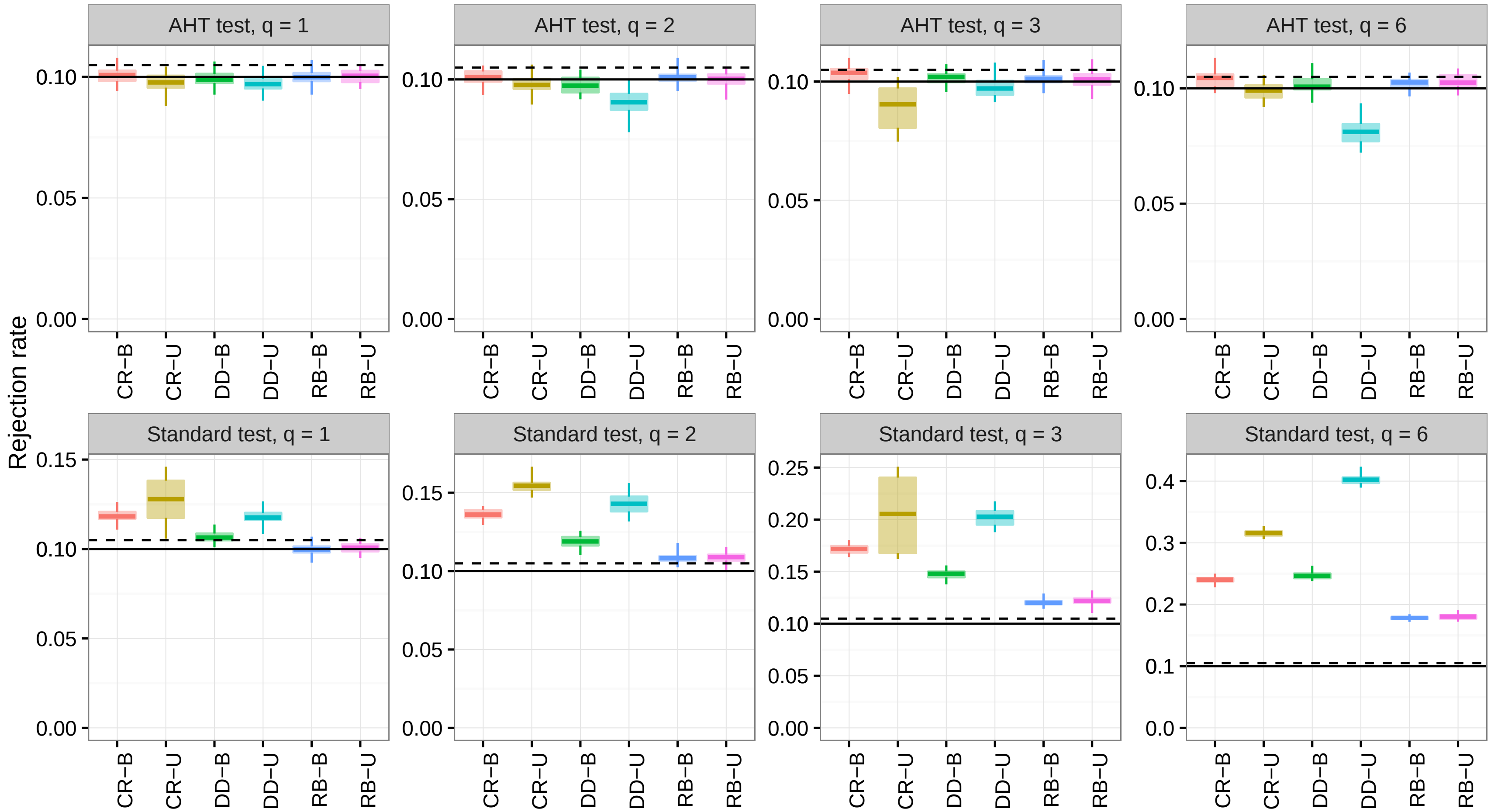

Figure S7: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis (q) for $\alpha = .10$ and $m = 30$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

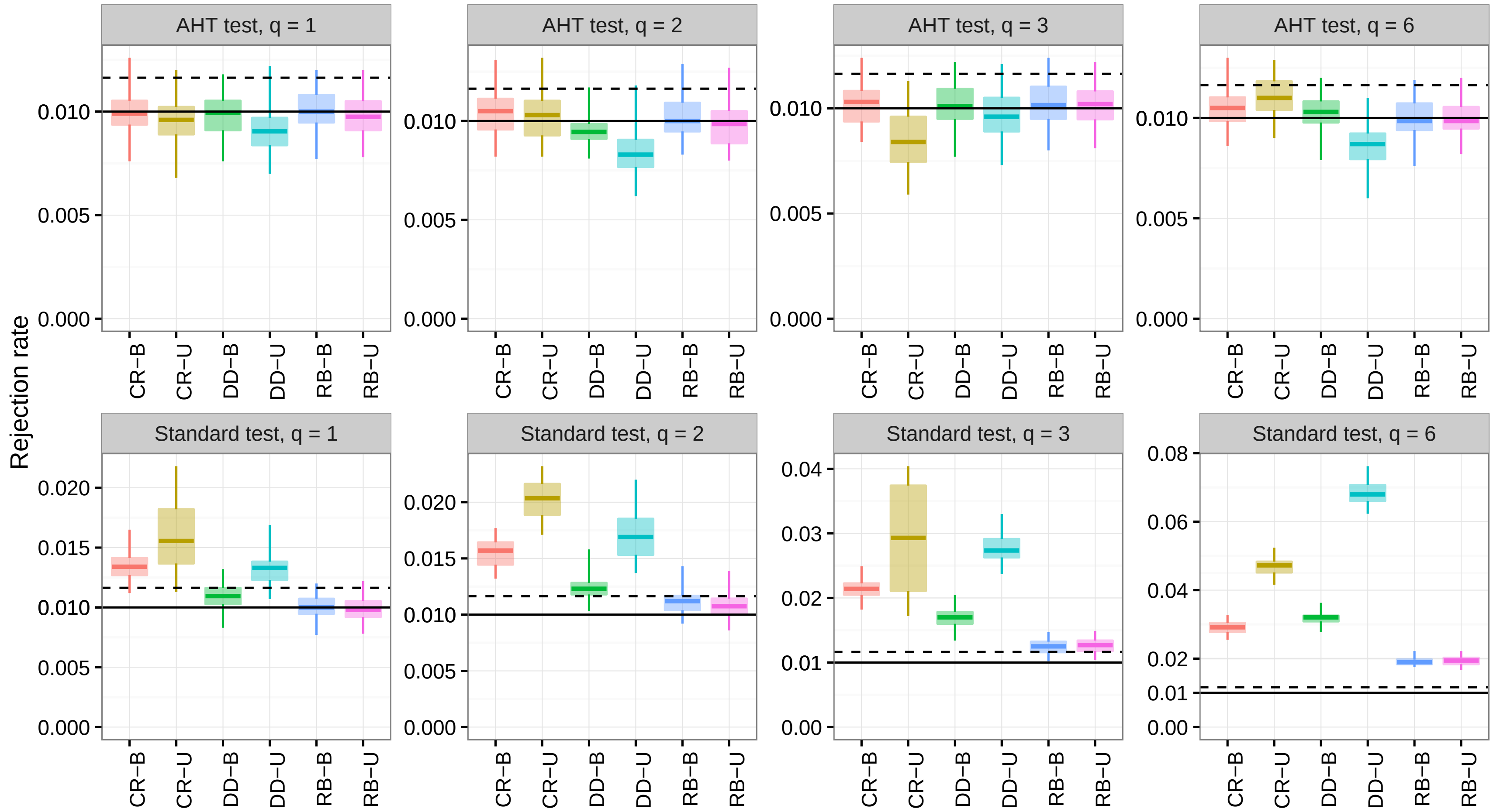

Figure S8: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis (q) for $\alpha = .01$ and $m = 50$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

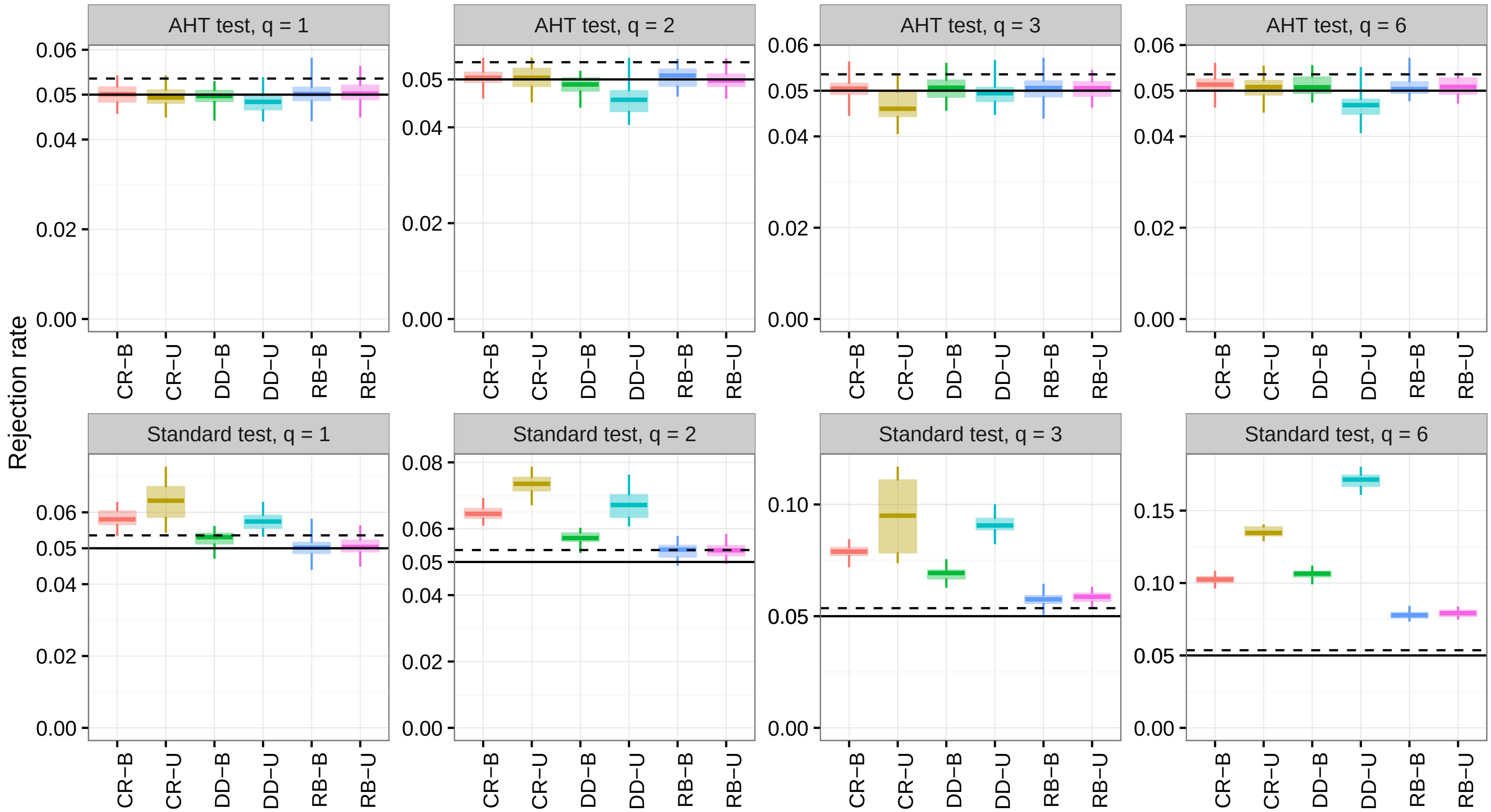

Figure S9: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis (q) for $\alpha = .05$ and $m = 50$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

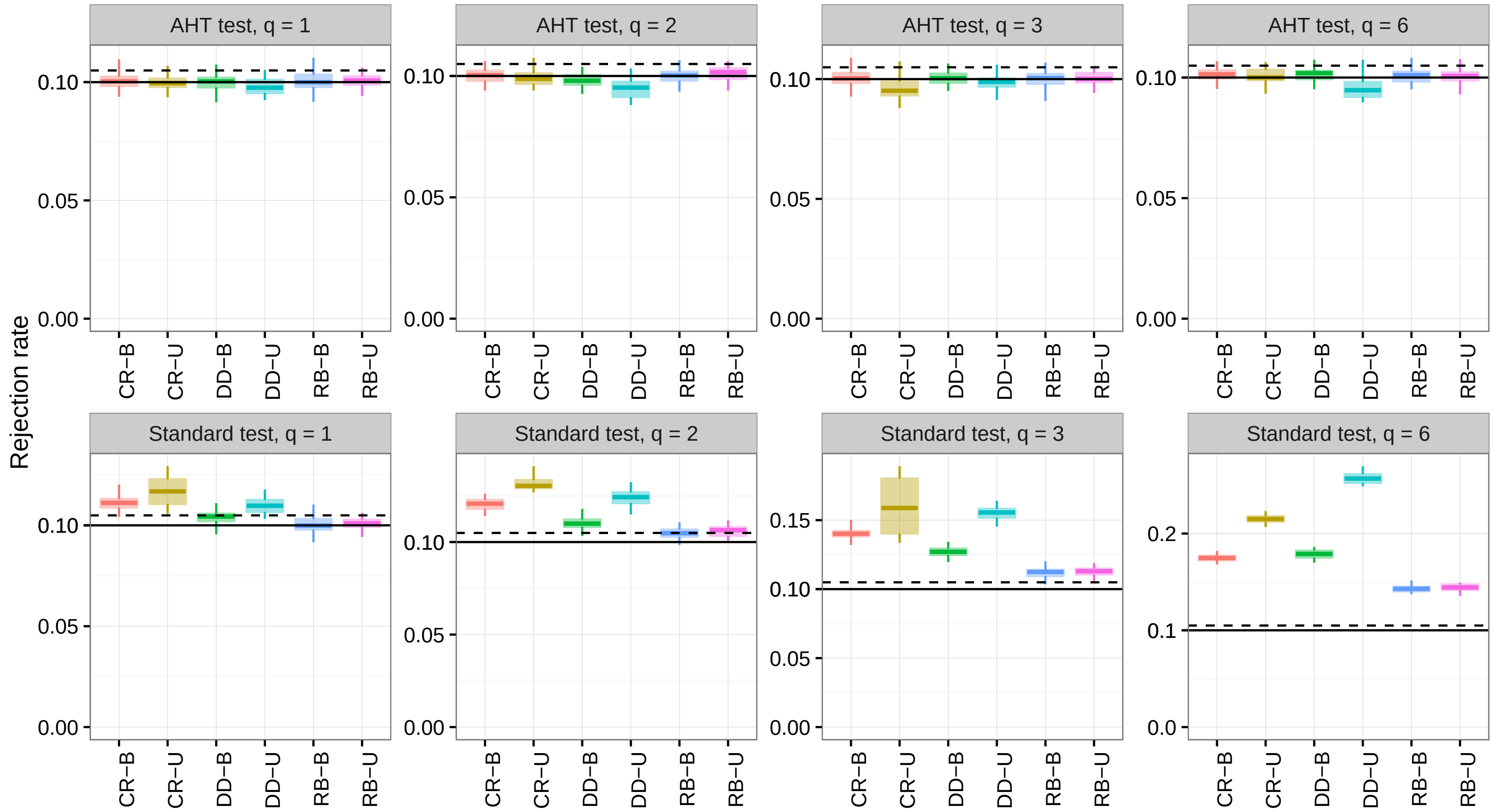

Figure S10: Rejection rates of AHT and standard tests, by study design and dimension of hypothesis (q) for $\alpha = .10$ and $m = 50$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

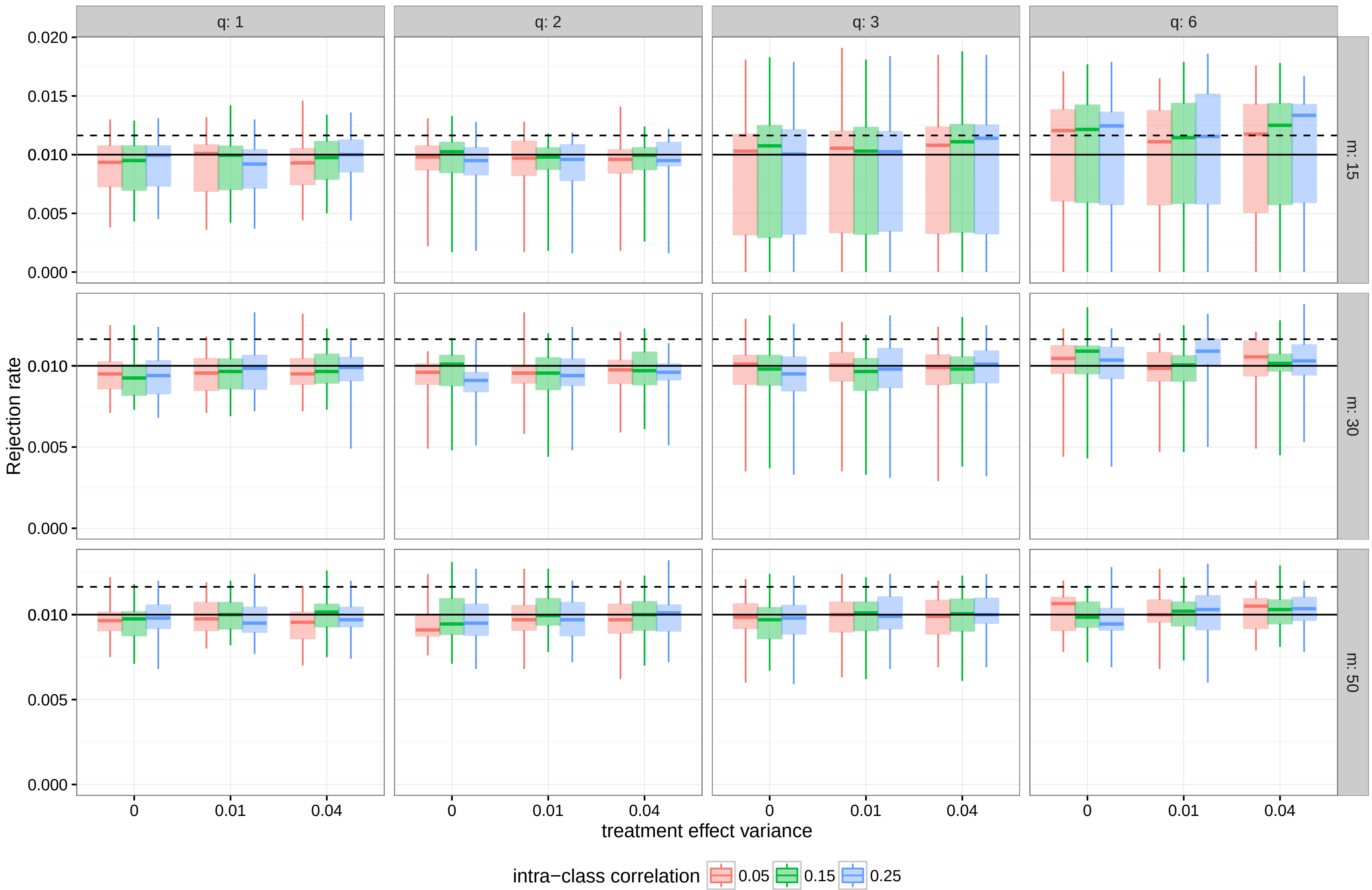

Figure S11: Rejection rates of AHT test, by treatment effect variance and intra-class correlation for $\alpha = .01$.

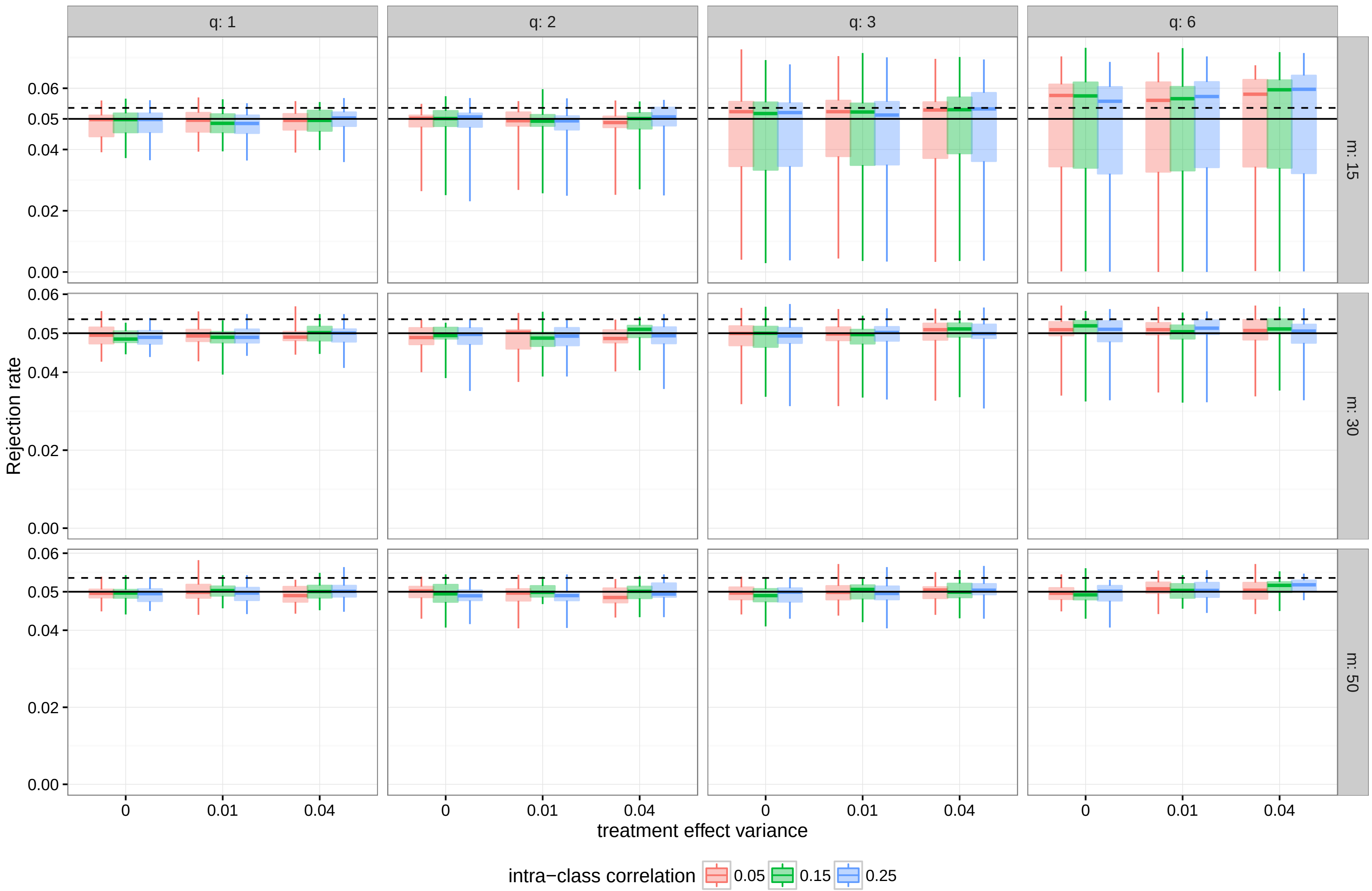

Figure S12: Rejection rates of AHT test, by treatment effect variance and intra-class correlation for $\alpha = .05$.

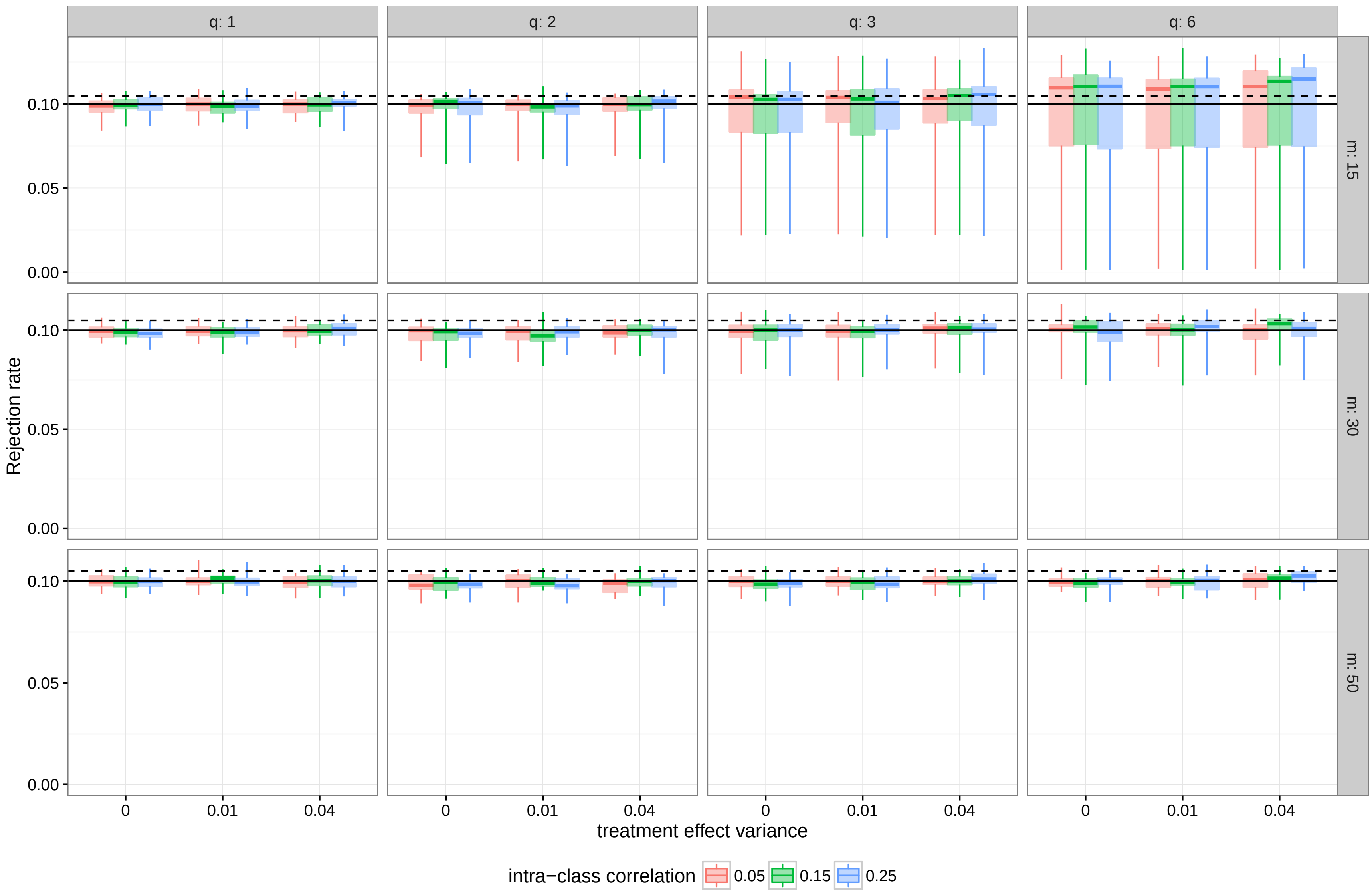

Figure S13: Rejection rates of AHT test, by treatment effect variance and intra-class correlation for $\alpha = .10$.

References

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