Supplementary Materials

Robust inference for geographic regression discontinuity designs: assessing the impact of police precincts

Emmett B. Kendall, Brenden Beck, Joseph Antonelli

Contents

1 Identification of causal effects

Here we show how the causal effects of interest can be identified from the observed data. First, we examine $\theta(R_\delta)$, which represents the average treatment effect within a distance of δ from the boundary.

$$\begin{aligned}
\theta(R_\delta) &= E[Y^1(R_\delta) - Y^0(R_\delta)] \\
&= E[Y^1(R_{\delta,1}) + Y^1(R_{\delta,0}) \\
&\qquad - Y^0(R_{\delta,1}) - Y^0(R_{\delta,0})] \\
&= E[Y^1(R_{\delta,1}) + Y^1(R_{\delta,1}) \\
&\qquad - Y^0(R_{\delta,0}) - Y^0(R_{\delta,0})] \quad \text{by assumption 2a} \\
&= 2E[Y^1(R_{\delta,1}) - Y^0(R_{\delta,0})] \\
&= 2E[Y(R_{\delta,1}) - Y(R_{\delta,0})] \quad \text{by assumption 1.}
\end{aligned}$$

Next, we examine identification of $\tau(\boldsymbol{b})$ for any location $\boldsymbol{b} \in \mathcal{B}$. Identification of τ follows immediately as it is simply a weighted average of $\tau(\boldsymbol{b})$ for some weight function $w(\boldsymbol{b})$. Note here that we use the notation $\lim_{\boldsymbol{s}\to\boldsymbol{b}^1}$ to denote a limit that approaches the boundary location $\boldsymbol{b}$ from the precinct 1 side of the boundary, with an analogous notation for precinct 0.

$$\begin{aligned}
\tau(\boldsymbol{b}) &= \lambda^1(\boldsymbol{b}) - \lambda^0(\boldsymbol{b}) \\
&= \lim_{\boldsymbol{s}\to\boldsymbol{b}^1} \lambda_1(\boldsymbol{b}) - \lim_{\boldsymbol{s}\to\boldsymbol{b}^0} \lambda_0(\boldsymbol{b}) \quad \text{by assumption 2b} \\
&= \lim_{\boldsymbol{s}\to\boldsymbol{b}^1} \lambda(\boldsymbol{b}) - \lim_{\boldsymbol{s}\to\boldsymbol{b}^0} \lambda(\boldsymbol{b}) \quad \text{by assumption 1.}
\end{aligned}$$

Both of these terms in the final calculation are identifiable from the observed data as we can estimate the observed intensity surface on the two sides of the boundary, separately.

2 Theoretical derivations

Here we provide the full mathematical details of the results shown in Section 3.5. First, we can show the type I error rate can be written as:

$$\begin{aligned}
P(\text{reject } H_0 \mid H_0) &= \int_q P(\text{reject } H_0 \mid H_0; \hat{Q}_{1-\alpha} = q) \cdot f_{\hat{Q}}(q) dq \\
&= \int_q P(Z_i > q; X_i) \cdot f_{\hat{Q}}(q) dq \\
&= \int_q [1 - P(Z_i \leq q; X_i)] \cdot f_{\hat{Q}}(q) dq \\
&= 1 - \int_q P(Z_i \leq q; X_i) \cdot f_{\hat{Q}}(q) dq \\
&= 1 - E_{\hat{Q}}[F(X_i, q)].
\end{aligned}$$

This shows that we need $E_{\widehat{Q}}\left[F(X_i, \widehat{Q}_{1-\alpha})\right] \geq 1 - \alpha$ in order to obtain type I error control. Next we highlight properties of our estimate of the CDF of the null distribution of the test statistic. First, we can show that the mean

of this estimate can be written as:

$$\begin{aligned}
E[\hat{F}(X_i,q)] &= E_{X^{(b)}}\Big[E\Big[\hat{F}(X_i,q)\Big|X_i^{(1)},X_i^{(2)},...,X_i^{(B)}\Big]\Big] \\
&= E_{X^{(b)}}\left[\frac{1}{B}\sum_{b=1}^{B}P(Z_i^{(b)}\leq q;X_i^{(b)})\right] \\
&= E_{X^{(b)}}\left[\frac{1}{B}\sum_{b=1}^{B}F(X_i^{(b)},q)\right] \\
&= E_{X^{(b)}}\Big[F(X_i^{(b)},q)\Big] \\
&\approx F(X_i,q)+\frac{d}{dX}F(X_i,q)\cdot E(X_i^{(b)}-X_i) \\
&\quad +\frac{d^2}{dX^2}F(X_i,q)\cdot E\Big[(X_i^{(b)}-X_i)^2\Big]
\end{aligned}$$

This shows that the mean of the CDF estimate depends on how far off the covariates in the null streets, $X_i^{(b)}$, are from the covariates at the precinct boundary of interest, denoted by X_i. Lastly, we can write the variance of our CDF estimate as:

$$\begin{aligned}
Var[\hat{F}(X_i,q)] &= E\Big[Var\Big[\hat{F}(X_i,q)\Big|X^{(b)}\Big]\Big] \\
&\quad +Var\Big[E\Big[\hat{F}(X_i,q)\Big|X^{(b)}\Big]\Big] \\
&= E\left[\frac{1}{B^2}\sum_{b=1}^{B}F(X_i^{(b)},q)\cdot(1-F(X_i^{(b)},q))\right] \\
&\quad +Var\left[\frac{1}{B}\sum_{b=1}^{B}F(X_i^{(b)},q)\right] \\
&= \frac{1}{B}E\Big[F(X_i^{(b)},q)\cdot(1-F(X_i^{(b)},q))\Big] \\
&\quad +\frac{1}{B}Var[F(X_i^{(b)},q)] \\
&\approx \frac{1}{B}\Bigg[E\Big[F(X_i^{(b)},q)\cdot(1-F(X_i^{(b)},q))\Big] \\
&\quad +Var\left[F(X_i;q)+\frac{d}{dX_i}F(X_i,q)(X_i^{(b)}-X_i)\right]\Bigg] \\
&= \frac{1}{B}\Bigg[E\Big(F(X_i^{(b)},q)\cdot(1-F(X_i^{(b)},q))\Big) \\
&\quad +\left(\frac{d}{dX_i}F(X_i,q)\right)^2\cdot Var(X_i^{(b)})\Bigg]
\end{aligned}$$

For simplicity of exposition, all of these results utilized a scalar covariate X_i, but could be easily extended to accommodate a vector of covariates to match on and analogous results would hold.

3 Investigating covariates to match on

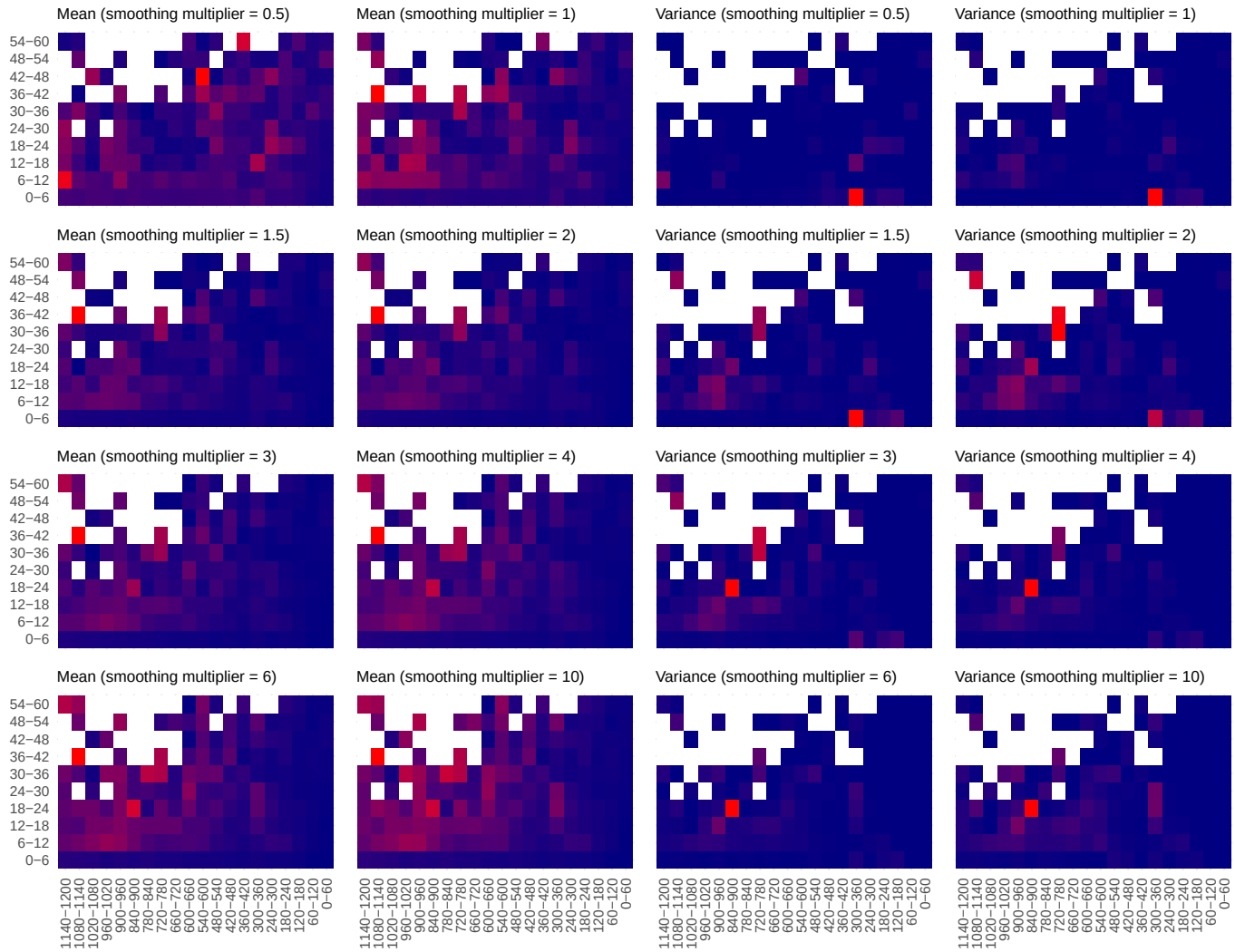


Figure 1: Heat map illustrating the mean (left half) and variance (right half) of the test statistics at the null streets as a function of the ratio of crime (y-axis) and the total amount of crime (x-axis).

As illustrated in the manuscript, it is important that we sample null streets with similar values of important covariates as the precinct boundaries of interest. This will help to ensure the test statistics at the null streets have a similar distribution as at the precinct borders under the null hypothesis. Figure 1 shows the mean and variance of the test statistics at the null streets after being binned according to the ratio of crime and total crime within a window around a null street. We see that there is a clear relationship between the ratio of crime or total crime and the mean and variance of the test statistic. In particular the mean of the test statistic depends heavily on both of these two variables. This motivates our procedure for matching on these variables when finding null streets for both the policing

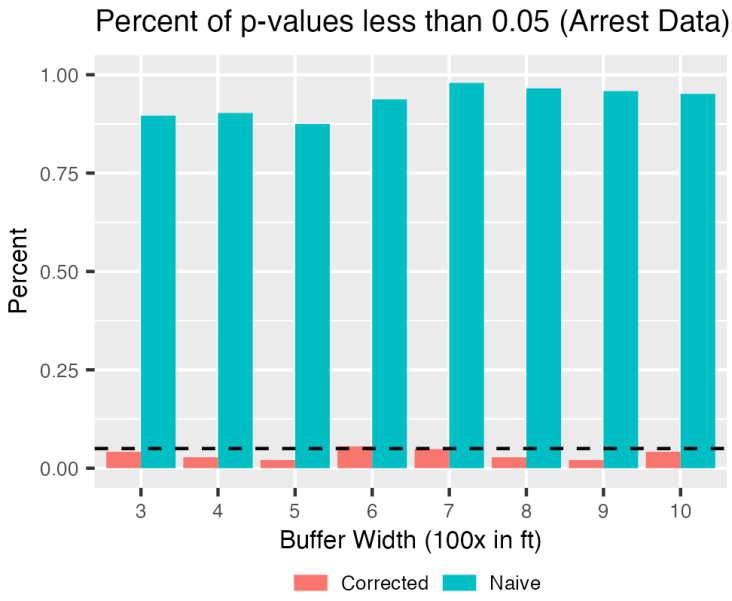
analysis and the negative control analysis.

4 Results using $\theta(R_\delta)$ as estimand

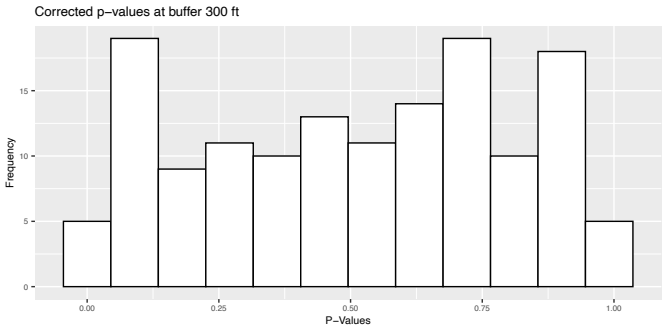
Note throughout this section that our outcomes $Y(R^{(i)}_{\delta,0})$ and $Y(R^{(i)}_{\delta,1})$ are the counts of the number of events within a buffer width of δ around boundary i, divided by a scaling factor dependent on the outcome being examined. For arrests, we divide the total number of arrests by the amount of crime in a region to examine arrest rates instead of counts. For the negative control outcome, we scale the number of trees by the total length of streets in the region to obtain a rate of trees per length of street. This is done to avoid issues caused by one side of the street having larger values of the outcome simply because they have more crime (or streets in the case of the negative control outcome).

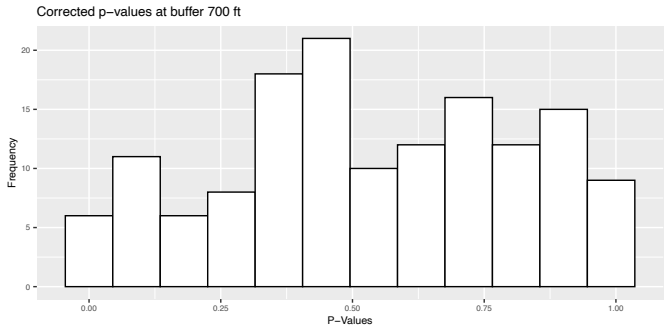
4.1 Arrest data

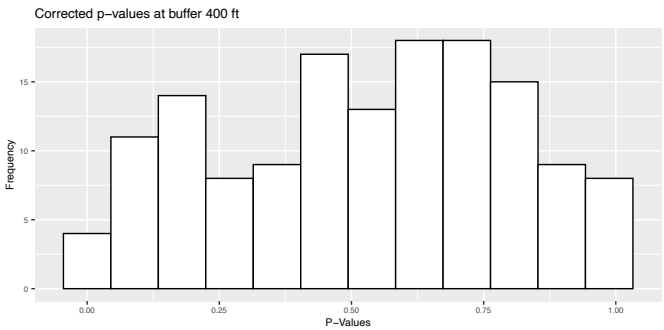
4.1.1 Proportion of p-values less than 0.05 using the naive versus corrected testing approach:

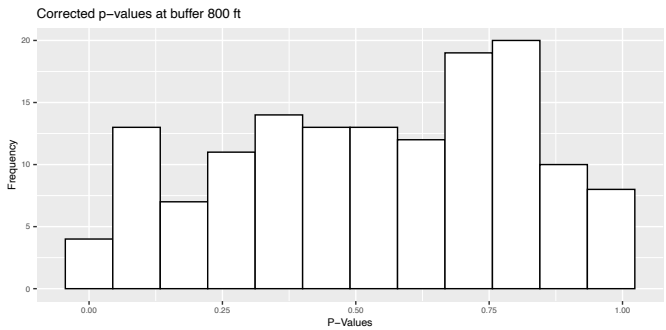


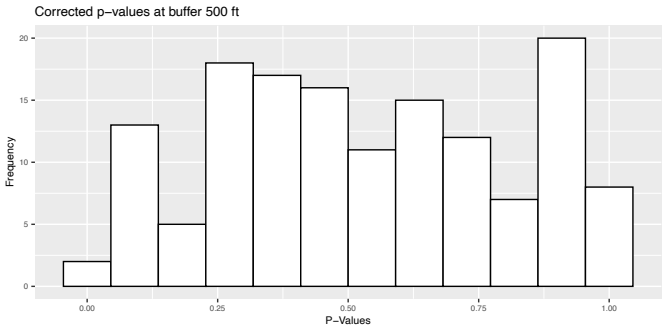
4.1.2 Histograms of the corrected p-values for differing buffer widths:

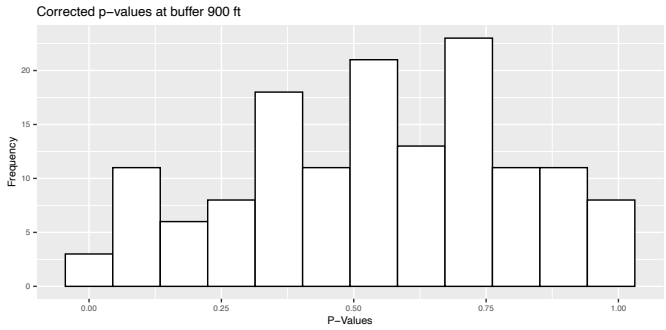


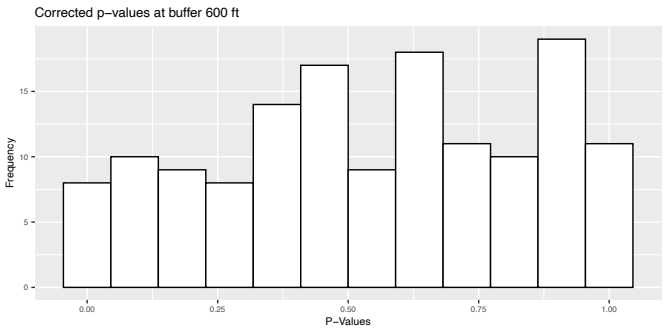


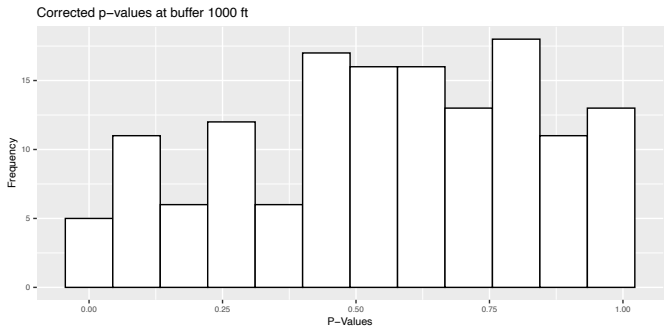




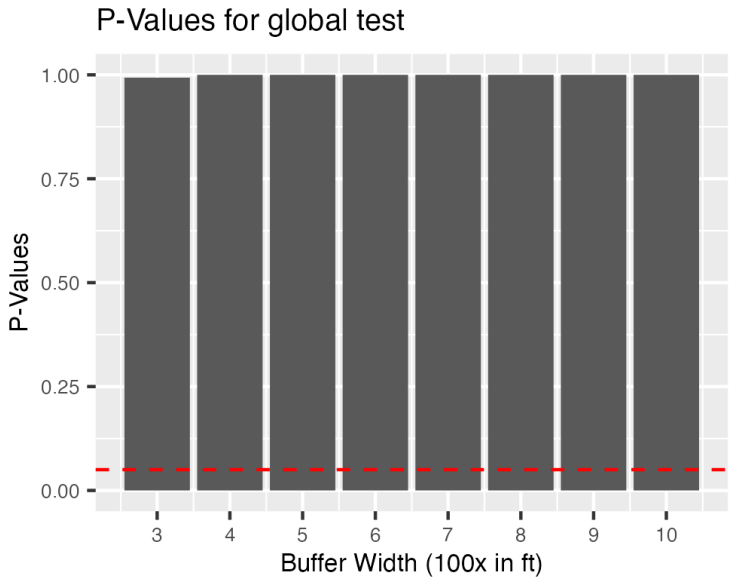






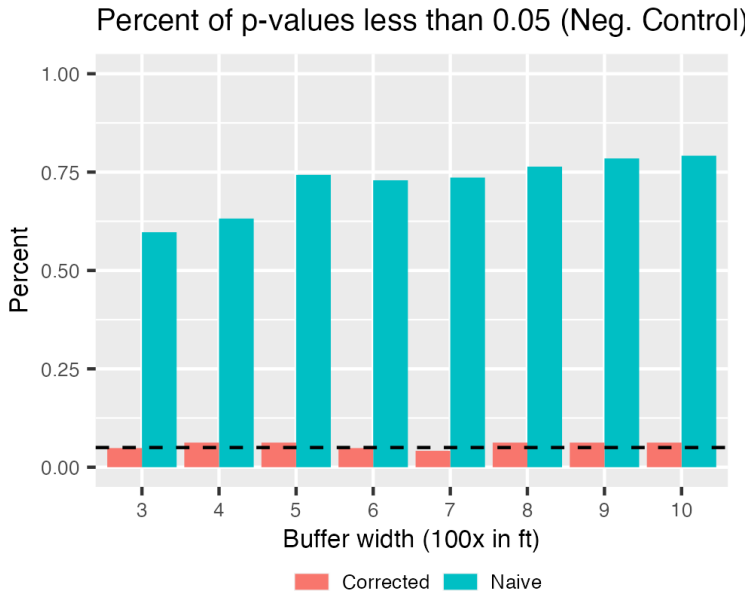


4.1.3 Global test results:

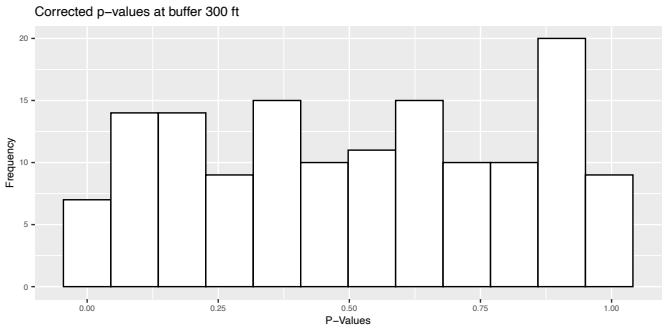


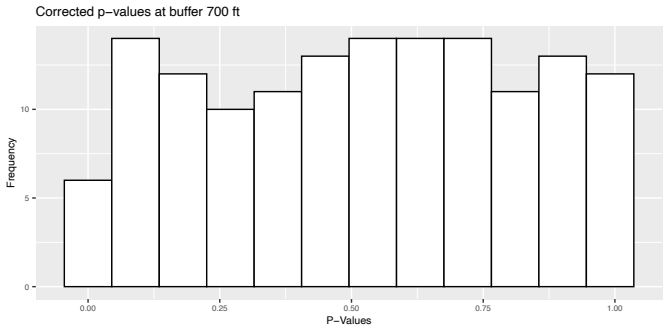
4.2 Negative control

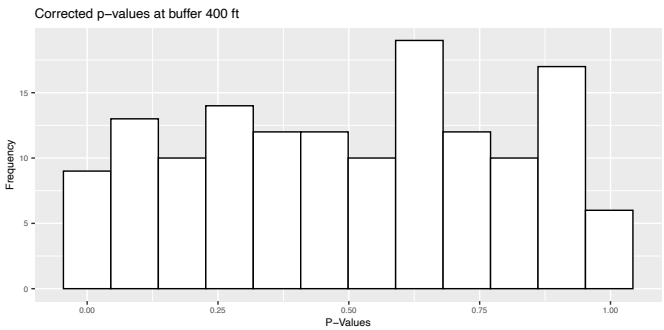
4.2.1 Proportion of p-values less than 0.05 using the naive versus corrected testing approach:

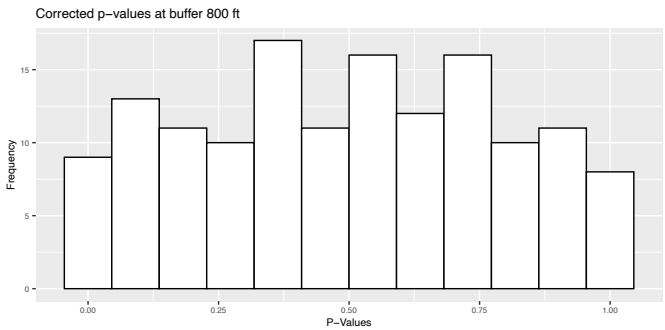


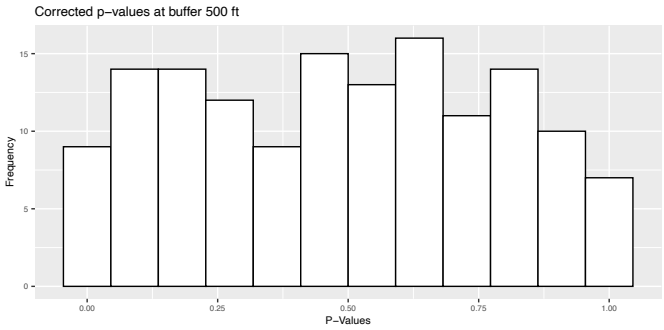
4.2.2 Histograms of the corrected p-values for differing buffer widths:

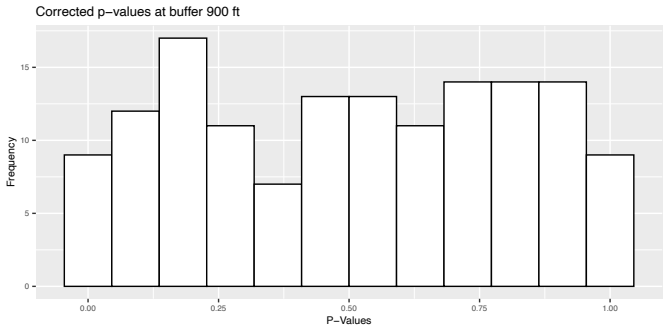


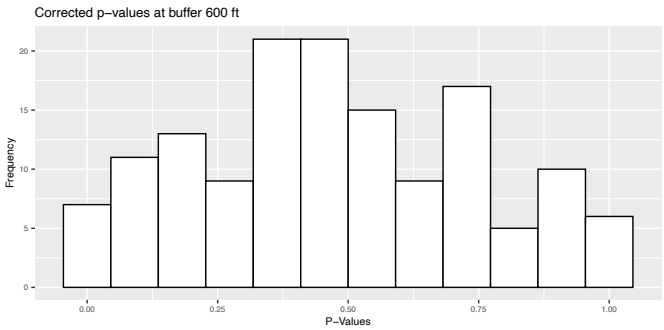


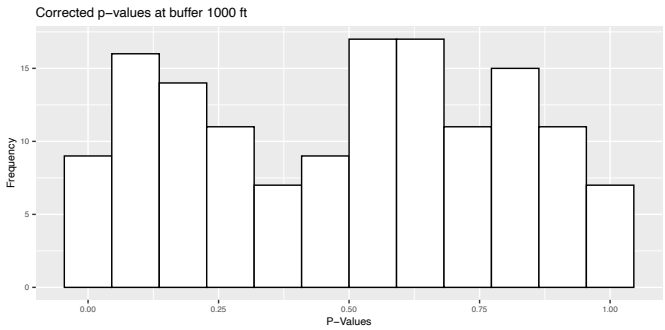




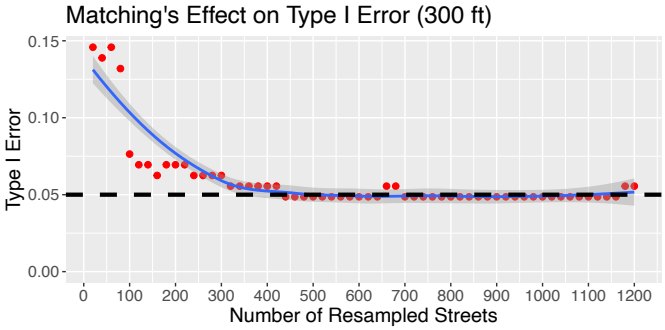


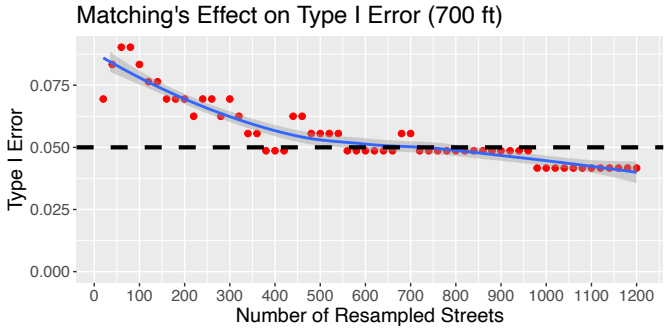


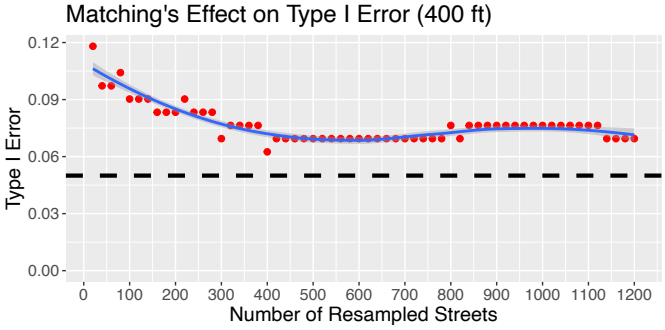


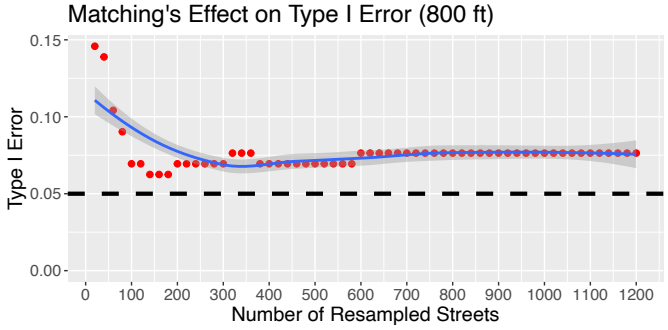


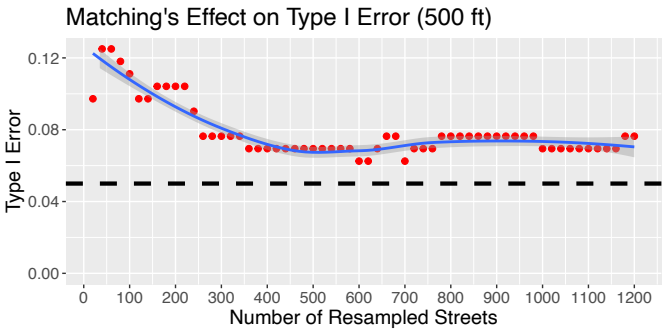
4.2.3 The relationship between matches and type I error probabilities for other buffer widths:

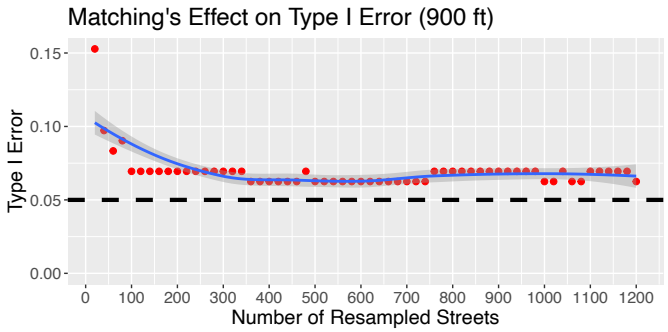


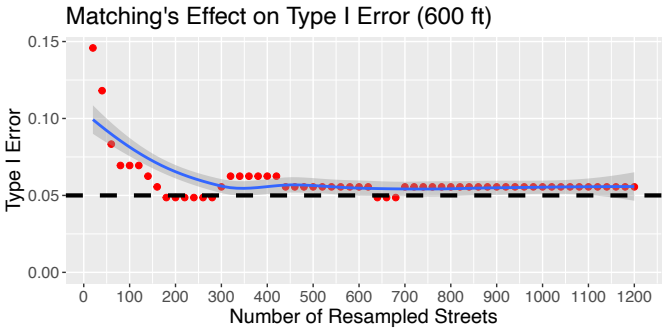


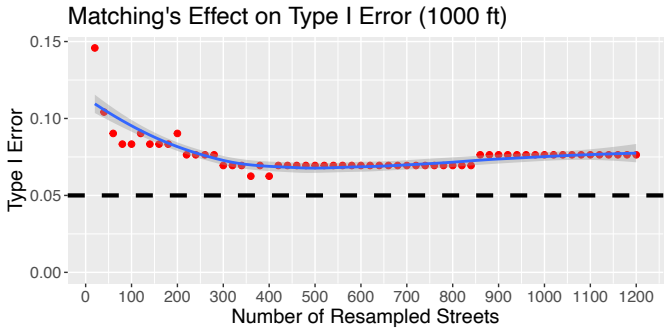








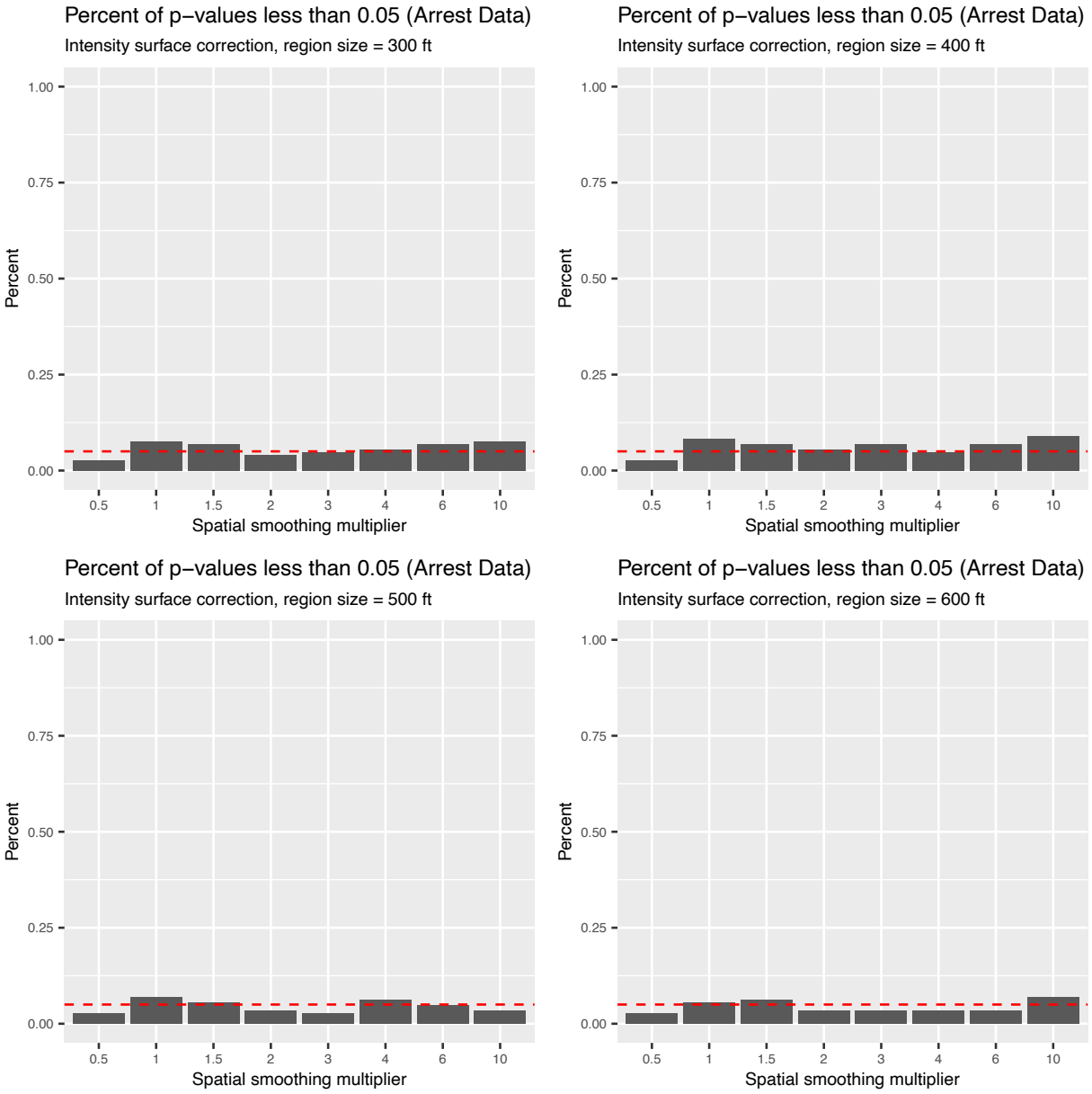


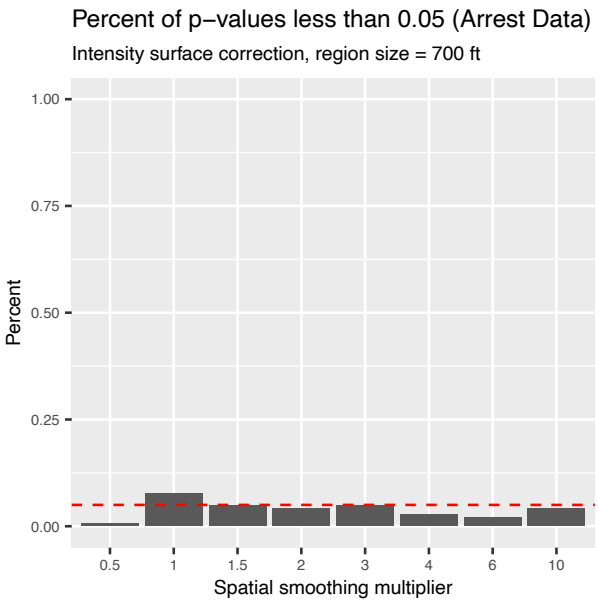


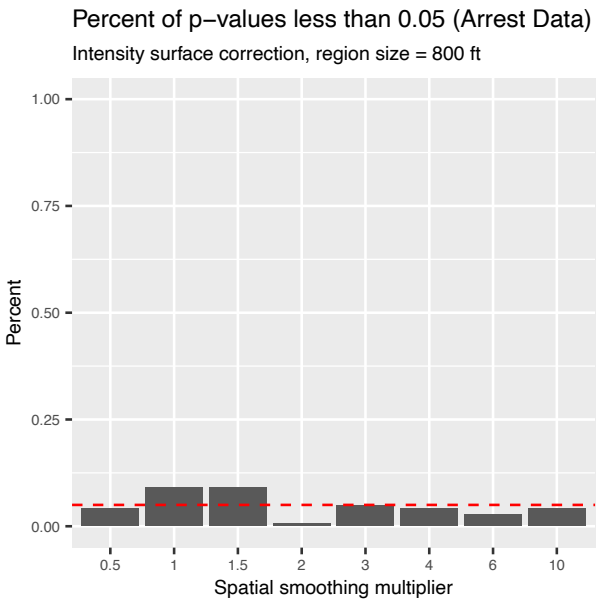
5 Additional results using $\tau(b)$ as estimand

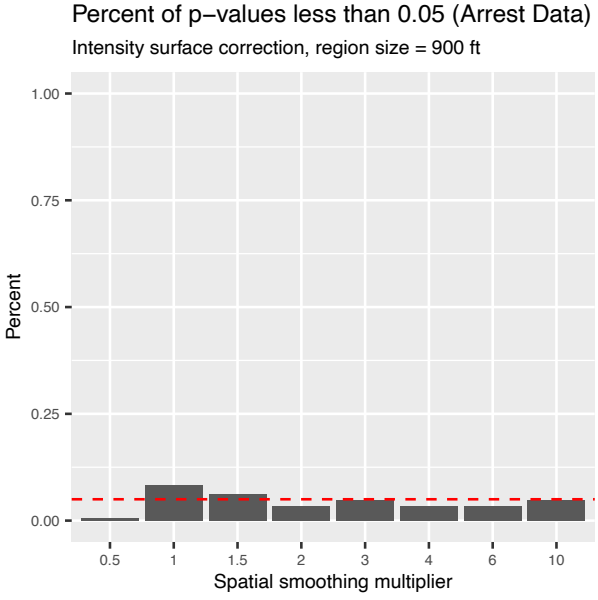
5.1 Arrest data

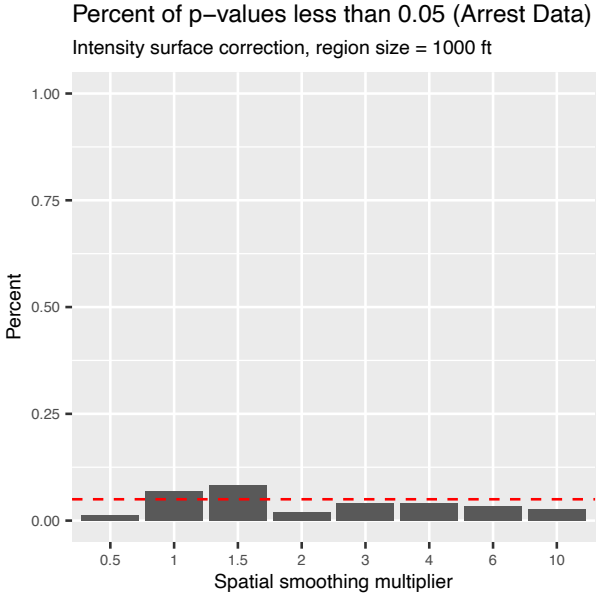
5.1.1 Type I error control across the different region sizes and levels of spatial smoothness:



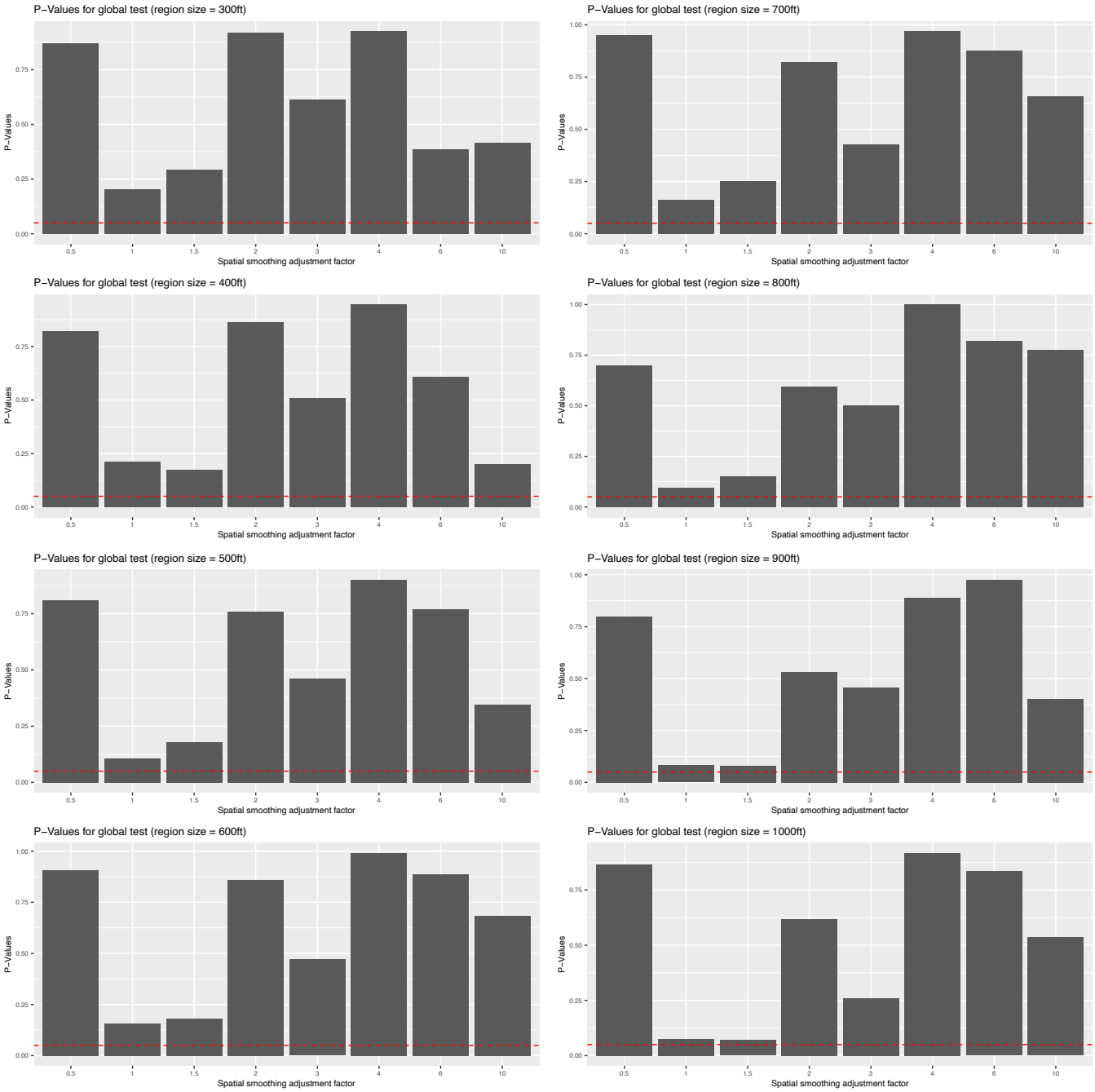

Percent of p−values less than 0.05 (Arrest Data)
Intensity surface correction, region size = 700 ft
Percent
1.00
0.75
0.50
0.25
0.00
0.5
1
1.5
2
3
4
6
10
Spatial smoothing multiplier


Percent of p−values less than 0.05 (Arrest Data)
Intensity surface correction, region size = 800 ft
Percent
1.00
0.75
0.50
0.25
0.00
0.5
1
1.5
2
3
4
6
10
Spatial smoothing multiplier


Percent of p−values less than 0.05 (Arrest Data)
Intensity surface correction, region size = 900 ft
Percent
1.00
0.75
0.50
0.25
0.00
0.5
1
1.5
2
3
4
6
10
Spatial smoothing multiplier

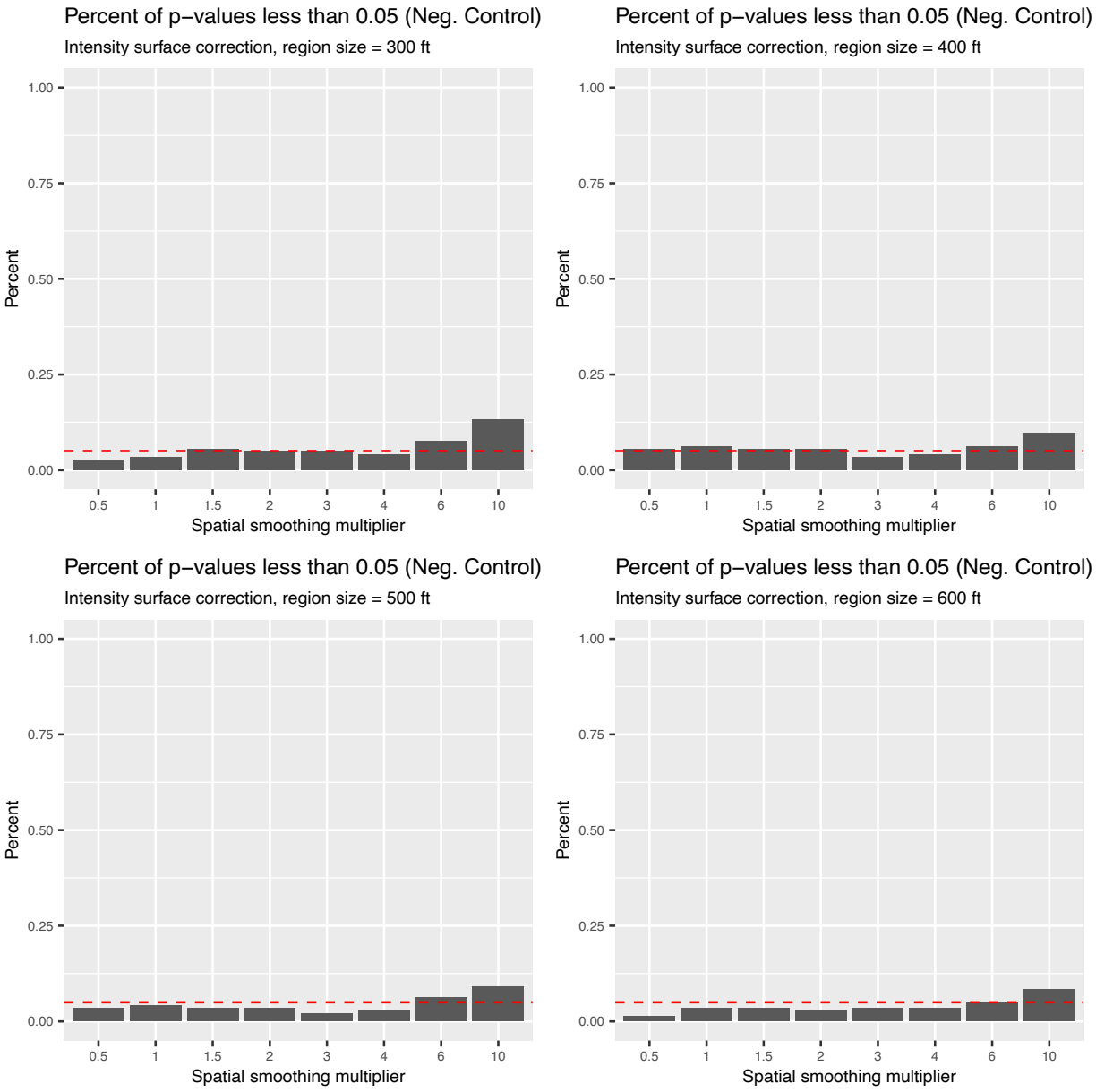

Percent of p−values less than 0.05 (Arrest Data)
Intensity surface correction, region size = 1000 ft
Percent
1.00
0.75
0.50
0.25
0.00
0.5
1
1.5
2
3
4
6
10
Spatial smoothing multiplier

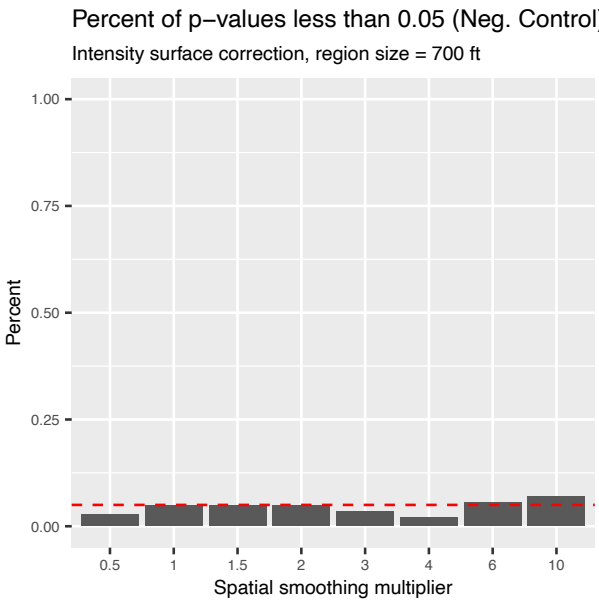
5.1.2 Global results across the different region sizes and levels of spatial smoothness:

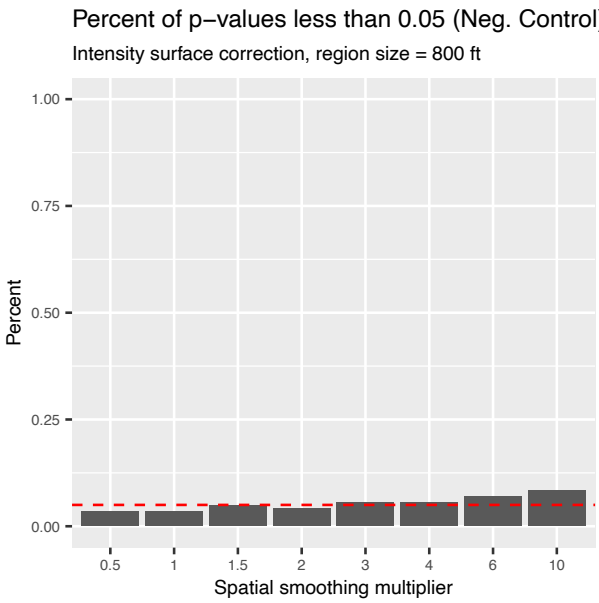


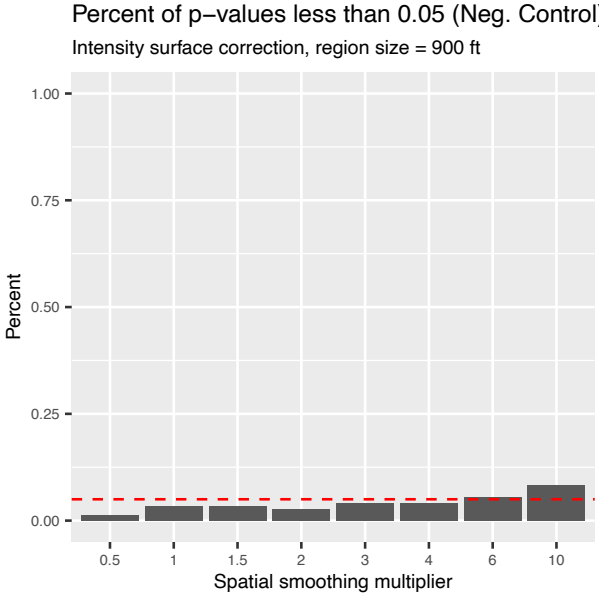
5.2 Negative control

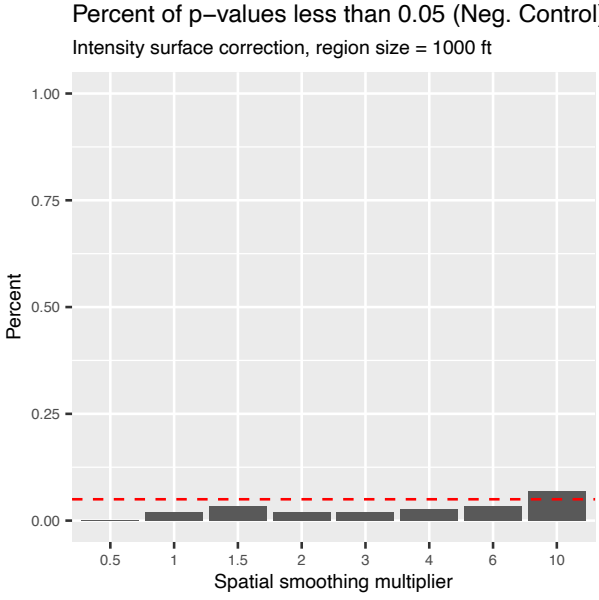
5.2.1 Type I error control across the different region sizes and levels of spatial smoothness:




Percent of p−values less than 0.05 (Neg. Control)
Intensity surface correction, region size = 700 ft
Percent
1.00
0.75
0.50
0.25
0.00
0.5
1
1.5
2
3
4
6
10
Spatial smoothing multiplier


Percent of p−values less than 0.05 (Neg. Control)
Intensity surface correction, region size = 800 ft
Percent
1.00
0.75
0.50
0.25
0.00
0.5
1
1.5
2
3
4
6
10
Spatial smoothing multiplier


Percent of p−values less than 0.05 (Neg. Control)
Intensity surface correction, region size = 900 ft
Percent
1.00
0.75
0.50
0.25
0.00
0.5
1
1.5
2
3
4
6
10
Spatial smoothing multiplier


Percent of p−values less than 0.05 (Neg. Control)
Intensity surface correction, region size = 1000 ft
Percent
1.00
0.75
0.50
0.25
0.00
0.5
1
1.5
2
3
4
6
10
Spatial smoothing multiplier