Supplementary Materials for
Modified Wilcoxon-Mann-Whitney Test and
Power against Strong Null

A Some details in the computation of $\mathrm{Var}(U)$

The variance of U equals

$$\mathrm{Var}\left\{\sum_{i=1}^{m}\sum_{j=1}^{n} I(X_i < Y_j)\right\} = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{i'=1}^{m}\sum_{j'=1}^{n} \mathrm{Cov}\left\{I(X_i < Y_j), I\left(X_{i'} < Y_{j'}\right)\right\}.$$

Of the $(mn)^2$ terms, there are four ways the indices i and j, i' and j' can overlap or be different. When $i \neq i'$ and $j \neq j'$, $I(X_i < Y_j)$ and $I\left(X_{i'} < Y_{j'}\right)$ are independent. Thus the total variance can be written as the sum of three parts:

$$\mathrm{Var}(U) = \frac{1}{(mn)^2}\sum_{i=1}^{m}\sum_{j\neq j'}^{n} \mathrm{Cov}\left\{I(X_i < Y_j), I\left(X_i < Y_{j'}\right)\right\} \tag{1}$$

$$+ \frac{1}{(mn)^2}\sum_{i\neq i'}^{m}\sum_{j=1}^{n} \mathrm{Cov}\left\{I(X_i < Y_j), I(X_{i'} < Y_j)\right\} \tag{2}$$

$$+ \frac{1}{(mn)^2}\sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{Var}\left\{I(X_i < Y_j)\right\}. \tag{3}$$

This expression can be made more explicit by noting that

$$\begin{aligned}
\mathrm{Cov}\left\{I(X_i < Y_j), I\left(X_i < Y_{j'}\right)\right\} &= \mathrm{E}\left\{I(X_i < Y_j) I\left(X_i < Y_{j'}\right)\right\} - \mathrm{E}\left\{I(X_i < Y_j)\right\}\mathrm{E}\left\{I\left(X_i < Y_{j'}\right)\right\} \\
&= \mathrm{E}\left[\mathrm{E}\left\{I(X_i < Y_j) I\left(X_i < Y_{j'}\right)|X_i\right\}\right] - \mathrm{E}\left[\mathrm{E}\left\{I(X_i < Y_j)|X_i\right\}\right]^2 \\
&= \mathrm{E}\left\{S_Y^2(X_i)\right\} - \mathrm{E}^2\left\{S_Y(X_i)\right\},
\end{aligned}$$

where $S_Y(X_i) = \mathrm{E}_{Y|X_i}\left\{I(X_i < Y)\right\}$. Thus

$$\begin{aligned}
\mathrm{Var}(U) &= (mn)^{-2}\left\{m\left(n^2 - n\right)\mathrm{Var}\left\{S_Y(X_i)\right\} + n\left(m^2 - m\right)\mathrm{Var}\left\{S_X(Y_i)\right\} + mn\,\mathrm{Var}\left\{I(X_i < Y_j)\right\}\right\} \\
&= (1 - 1/n)\mathrm{Var}\left\{S_Y(X_i)\right\}/m + (1 - 1/m)\mathrm{Var}\left\{S_X(Y_i)\right\}/n + \mathrm{Var}\left\{I(X_i < Y_j)\right\}/(mn) \\
&= (1 - 1/n)\mathrm{Var}\left\{F_Y(X_i)\right\}/m + (1 - 1/m)\mathrm{Var}\left\{F_X(Y_i)\right\}/n + \mathrm{Var}\left\{I(X_i < Y_j)\right\}/(mn),
\end{aligned}$$

where $S_X(Y_i) = \mathrm{E}_{X|Y_i}\left\{I(Y_i < X)\right\}$. Under the strong null, $F_Y(X_i)$ and $F_X(Y_i)$ are uniform random variables, and we can verify that this expression evaluates to

$$\begin{aligned}
&= (mn)^{-2}\left\{m\left(n^2 - n\right)/12 + n\left(m^2 - m\right)/12 + mn/4\right\} \\
&= (mn)^{-2}\left(mn^2/12 + m^2n/12 + mn/12\right) \\
&= \left\{1/m + 1/n + 1/(mn)\right\}/12.
\end{aligned}$$

We can estimate $\mathrm{Var}(U)$ by

$$(1 - 1/n)\widehat{\mathrm{Var}}\left\{\hat{F}_Y(X_i)\right\}/m + (1 - 1/m)\widehat{\mathrm{Var}}\left\{\hat{F}_X(Y_i)\right\}/n + \widehat{\mathrm{Var}}\left\{I(X_i < Y_j)\right\}/(mn), \tag{4}$$

where $\widehat{\text{Var}}\left\{I\left(X_i<Y_j\right)\right\}$ is $\mathbb{P}_m\left\{\hat{F}_Y\left(X_i\right)\right\}\mathbb{P}_n\left\{\hat{F}_X\left(Y_i\right)\right\}$ and $\mathbb{P}_m$ is the empirical average operator corresponding to the m observations $X_1,\cdots,X_m$ and $\mathbb{P}_n$ is the empirical average operator corresponding to the n observations $Y_1,\cdots,Y_n$.

Fligner and Policello (1981) proposed to estimate $\text{Var}\left(U\right)$ by

$$
\begin{aligned}
&(m-1)\,n^2\widehat{\text{Var}}\left\{\hat{F}_Y\left(X_i\right)\right\}/\left(mn\right)^2+\left(n-1\right)m^2\widehat{\text{Var}}\left\{\hat{F}_X\left(Y_i\right)\right\}/\left(mn\right)^2+mn\mathbb{P}_m\left\{\hat{F}_Y\left(X_i\right)\right\}\mathbb{P}_n\left\{\hat{F}_X\left(Y_i\right)\right\}/\left(mn\right)^2\\
&=\left(1-1/m\right)\widehat{\text{Var}}\left\{\hat{F}_Y\left(X_i\right)\right\}/m+\left(1-1/n\right)\widehat{\text{Var}}\left\{\hat{F}_X\left(Y_i\right)\right\}/n+\mathbb{P}_m\left\{\hat{F}_Y\left(X_i\right)\right\}\mathbb{P}_n\left\{\hat{F}_X\left(Y_i\right)\right\}/\left(mn\right)\\
&=\left(1-1/m\right)\widehat{\text{Var}}\left\{\hat{F}_Y\left(X_i\right)\right\}/m+\left(1-1/n\right)\widehat{\text{Var}}\left\{\hat{F}_X\left(Y_i\right)\right\}/n+\widehat{\text{Var}}\left\{I\left(X_i<Y_j\right)\right\}/\left(mn\right).
\end{aligned}
$$

This estimate is slightly different from (4). It is not as good an estimate because, for example, under the strong null it is an estimate of

$$
\begin{aligned}
&=\left(1-1/m\right)/m/12+\left(1-1/n\right)/n/12+1/\left(mn\right)/4\\
&=\left\{1/m+1/n+\left(3/mn-n^{-2}-m^{-2}\right)\right\}/12,
\end{aligned}
$$

which does not equal $\left\{1/m+1/n+1/\left(mn\right)\right\}/12$ when $m\neq n$. The differences are not important though because it is on the order of $O\left(m^{-1}n^{-1}\right)$.

B Proof of Theorem 1

Proof. To show that $\min\left(\hat{V},V_{mn}\right)-V_{mn}$ converges to 0 in probability, it suffices to show that $\min\left(\hat{V}-V_{mn},0\right)\to_p 0$. Because $\left|\min\left(\hat{V}-V_{mn},0\right)\right|<\left|\hat{V}-V_{mn}\right|$, this, in turn, holds if $\left|\hat{V}-V_{mn}\right|\to_p 0$. Because $\hat{V}$ is a consistent estimator of V_{mn} and the function $|\cdot|$ is a continuous function, by the continuous mapping theorems for convergence in probability, we have $\left|\hat{V}-V_{mn}\right|\to_p 0$.

By the central limit theorem for U-statistics (Lee, 1990), we know $\left(U-1/2\right)/\sqrt{V_{mn}}\to_d N\left(0,1\right)$. Thus, according to Slutsky's theorem and $\min\left(\hat{V},V_{mn}\right)\to_p V_{mn}$, it follows that

$$
\frac{U-1/2}{\sqrt{\min\left(\hat{V},V_{mn}\right)}}\to_d N\left(0,1\right).
$$

■

C Ties

When ties exist, modifications of the U statistic and variance estimates are required.

$$
U=\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\phi\left(X_i,Y_j\right),
$$

where $\phi\left(X_i,Y_j\right)=1$ if $X_i<Y_j$, 1/2 if $X_i=Y_j$, and 0 if $X_i>Y_j$. For the rank sum representation, it is equivalent to "giving tied observations the average of the ranks for which those observations are competing" (Hollander et al., 2013, p. 118 and 128).

Under the strong null, the variance of U is

$$V_{mn} = \frac{1}{12}\left(\frac{1}{m} + \frac{1}{n} + \frac{1}{mn}\right) - \frac{1}{12}\frac{\sum_{j=1}^{g}(t_j - 1)\,t_j\,(t_j + 1)}{mnN\,(N-1)},$$

where g is the number of tied groups and t_j is the size of tied group j.

The variance estimate for U in the FP test can be adjusted through adjustment to, e.g. $\hat{G}$. Instead of defining it as the empirical distribution function estimated from $\{Y_j\}_{j=1}^{n}$, we will have (Hollander et al., 2013, p. 146)

$$\hat{G}\,(X) = \frac{1}{n}\sum_{j=1}^{n}\phi\,(X, Y_j)\,.$$

Permutation distribution can be obtained as when there are no ties.

D Additional Monte Carlo results

D.1 Monte Carlo results using normal distributions

Table D.1 and Figure D.1 repeat the simulation studies of Section 2 using normal instead of logistic distributions.

m	$n = m$			$n = 2m$			$n = 4m$		
	FP*	(2)	(3)	FP*	(2)	(3)	FP*	(2)	(3)
$\sigma_y = 1$									
5	10.5	10.5	9.1	5.3	8.3	6.0	-3.3	5.3	3.4
10	7.7	7.7	4.8	3.9	5.5	3.2	-1.0	3.4	2.0
20	4.4	4.4	2.5	2.4	3.2	1.8	-0.2	2.0	1.2
30	3.1	3.1	1.6	1.6	2.2	1.2	-0.1	1.4	0.8
50	1.9	1.9	1.0	1.0	1.3	0.7	-0.1	0.8	0.4
$\sigma_y = 2$									
5	7.8	7.8	7.6	12.4	8.6	6.3	8.3	7.2	4.6
10	5.8	5.8	3.8	7.4	5.5	3.2	4.9	4.4	2.5
20	3.4	3.4	1.9	4.0	3.0	1.6	2.7	2.5	1.3
30	2.3	2.3	1.2	2.8	2.1	1.1	1.8	1.7	0.9
50	1.5	1.5	0.8	1.7	1.3	0.7	1.1	1.0	0.5
$\sigma_y = 0.5$									
5	7.5	7.5	7.1	-3.1	4.6	3.7	-10.8	2.2	1.5
10	5.8	5.8	3.8	-0.6	3.3	2.0	-4.9	1.7	1.0
20	3.5	3.5	2.1	0.1	2.1	1.3	-2.1	1.2	0.8
30	2.4	2.4	1.3	0.1	1.4	0.8	-1.4	0.8	0.5
50	1.5	1.5	0.8	0.1	0.9	0.5	-0.9	0.5	0.3

Table D.1: Percent bias of three variance estimators. FP*: the original variance estimator from Fligner and Policello (1981); (2) and (3): variance estimators based on formula (2) and (3). $X \sim N(0,1), Y \sim N(0, \sigma_y^2)$.

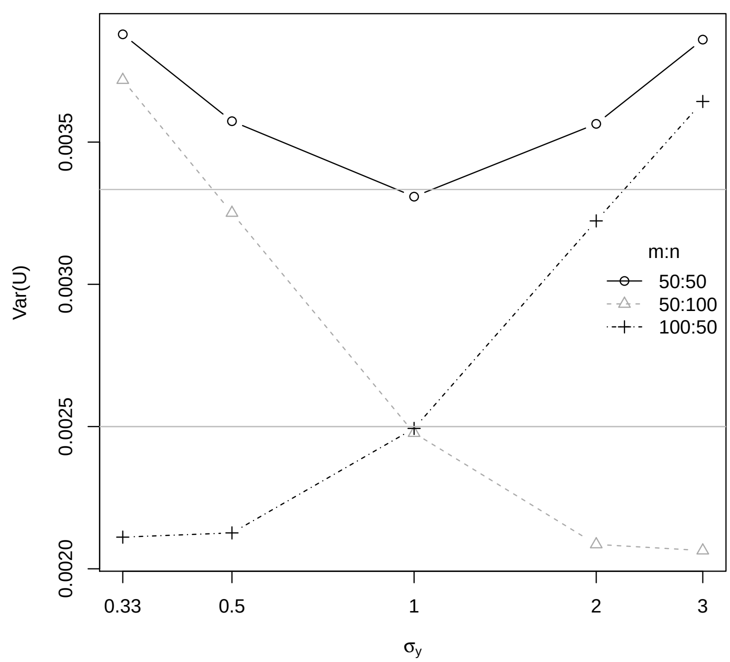

Figure D.1: Relationship between Var(U) and s_y. $X \sim N(0,1)$, $Y \sim N(0, \sigma_y^2)$. The horizontal lines have heights V_{mn}, the theoretical variance of U when $\sigma_y = 1$.

Table D.2 and D.3 repeat the simulation studies from Section 3 using normal instead of logistic distributions.

m	$n = m$				$n = 2m$				$n = 4m$			
	WMW	FP*	(2)	(3)	WMW	FP*	(2)	(3)	WMW	FP*	(2)	(3)
normal approx.												
5	3.32	8.03	8.03	8.03	4.19	9.05	8.89	9.14	4.16	10.13	9.69	10.07
10	4.36	6.54	6.54	6.83	4.44	6.79	6.69	6.92	4.95	7.53	7.11	7.29
20	4.68	5.32	5.32	5.46	4.72	5.53	5.45	5.61	4.57	5.58	5.43	5.51
30	4.85	5.52	5.52	5.68	4.78	5.40	5.33	5.45	5.05	5.79	5.60	5.69
50	5.02	5.22	5.22	5.31	4.52	4.95	4.90	4.97	4.96	5.41	5.37	5.38
permutation												
5	3.32	4.17	4.17	4.17	4.19	5.21	5.32	5.25	4.16	5.23	5.21	5.22
10	4.36	5.08	5.08	5.10	4.94	5.09	5.18	5.18	5.21	5.11	5.10	5.12
20	4.68	4.80	4.80	4.79	4.72	4.74	4.75	4.73	4.57	4.66	4.67	4.66
30	4.97	5.05	5.05	5.05	4.78	4.85	4.85	4.85	5.05	5.00	5.04	5.05
50	5.02	5.06	5.06	5.05	4.55	4.59	4.60	4.59	4.96	4.96	4.98	4.97

Table D.2: Estimated size (%) of three FP tests based on different variance estimators, nominal level 5%. FP*: the original variance estimator from Fligner and Policello (1981); (2) and (3): variance estimators based on formula (2) and (3). $X, Y \sim N(0, 1)$.

σ_y	$n = m$		$n = 2m$		$n = 4m$	
	WMW	FP	WMW	FP	WMW	FP
normal approx., $m = 30$						
1	45	47	58	59	65	67
2	29	28	36	44	41	60
0.5	46	46	58	50	64	52
permutation, $m = 30$						
1	45	46	58	58	65	65
2	29	27	36	42	41	58
0.5	47	45	58	48	64	50
permutation, $m = 10$						
1	51	54	68	67	76	74
2	33	34	44	49	52	66
0.5	52	52	67	58	73	59

Table D.3: Estimated power (%) of the WMW and FP tests. The FP test uses the variance estimator based on formula (2). For the WMW test, p-value determined by normal approximation when either m or n exceeds 50 and by exact distribution otherwise; for the FP test, p-value determined according to the method indicated. $X \sim N(0, 1)$, $Y \sim N(\mu_y, \sigma_y^2)$. μ_y is set to 0.5, 0.6 and 0.4, respectively for the three σ_y at $m = 30$ and 1, 1.2 and 0.8, respectively for the three σ_y at $m = 10$ to achieve comparable level of power.

Table D.4 and D.5 repeat the simulation studies from Section 4 using normal instead of logistic distributions.

m	$n=m$		$n=2m$		$n=4m$	
	FP	C	FP	C	FP	C
normal approx.						
5	8.03	8.03	8.89	8.89	9.69	9.79
10	6.54	6.54	6.69	6.72	7.11	7.28
20	5.32	5.33	5.45	5.62	5.43	5.67
30	5.52	5.52	5.33	5.48	5.60	5.83
50	5.22	5.23	4.90	4.99	5.37	5.54
permutation						
5	4.17	4.17	5.32	5.32	5.21	5.21
10	5.08	5.08	5.18	5.14	5.10	5.15
20	4.80	4.81	4.75	4.80	4.67	4.62
30	5.05	5.07	4.85	4.87	5.04	5.04
50	5.06	5.07	4.60	4.61	4.98	5.00

Table D.4: Estimated size (%) of the FP and combined (C) tests, nominal level 5%. $X, Y \sim N(0,1)$.

σ_y	$n=m$				$n=2m$				$n=4m$			
	vdW	W	FP	C	vdW	W	FP	C	vdW	W	FP	C
normal approx., $m=30$												
1	46	45	47	47	59	58	59	60	67	65	67	68
2	25	29	28	29	26	36	44	44	27	41	60	60
0.5	42	46	46	47	58	58	50	58	68	64	52	64
permutation, $m=30$												
1	46	45	46	46	59	58	58	58	67	65	65	65
2	26	29	27	28	26	36	42	42	27	41	58	58
0.5	43	47	45	45	58	58	48	56	68	64	50	62
permutation, $m=10$												
1	55	51	54	54	69	68	67	68	78	76	74	74
2	33	33	34	34	37	44	49	49	40	52	66	65
0.5	52	52	52	52	67	67	58	61	77	73	59	66

Table D.5: Estimated power (%) of the van der Waerden (vdW), WMW (W), FP, and the combined (C) tests. $X \sim N(0,1)$, $Y \sim N(\mu_y, \sigma_y^2)$. μ_y is set to 0.5, 0.6 and 0.4, respectively for the three σ_y at $m=30$ and 1, 1.2 and 0.8, respectively for the three σ_y at $m=10$ to achieve comparable level of power.

D.2 Power studies using other distributions

In addition to the power comparison using the logistic distributions (Table 5) and the normal distributions (Table D.5), here we look at two more examples where the distributions may not be symmetric and the shape of the distributions in the two samples are different.

Figure D.2 shows the density functions of *lognormal* $(0, \sigma = 1)$ and *lognormal* $(0, \sigma = 0.25)$+0.4. Table D.6 shows the result of the power study from Section 4 using these two distributions.

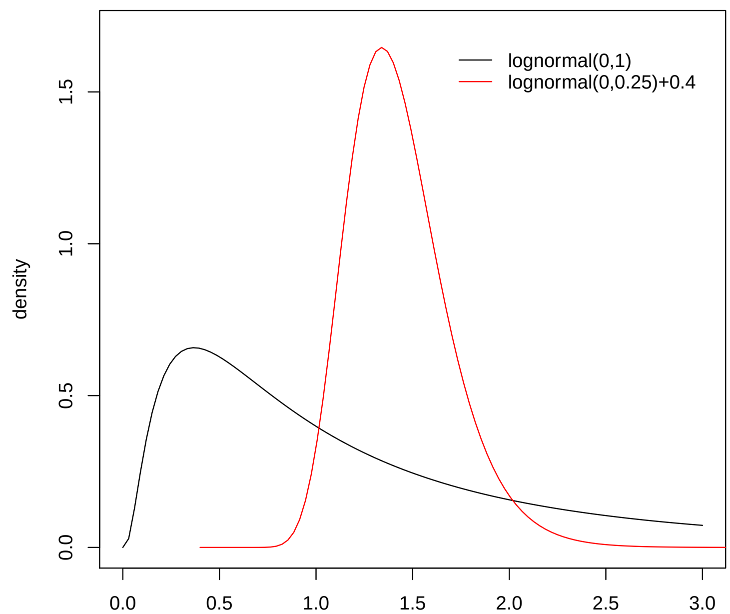

Figure D.2: Densities of two lognormal distributions.

	$n = m$				$n = 2m$				$n = 4m$			
	vdW	W	FP	C	vdW	W	FP	C	vdW	W	FP	C
normal approx., $m = 30$												
$X \sim LN(0,1)$, $Y \sim LN(0,0.25) + 0.4$	31	43	36	43	48	53	37	53	59	59	37	59
$X \sim LN(0,0.25) + 0.4$, $Y \sim LN(0,1)$	31	43	36	43	35	54	61	61	36	65	87	87
permutation, $m = 30$												
$X \sim LN(0,1)$, $Y \sim LN(0,0.25) + 0.4$	32	43	35	41	48	53	36	52	59	59	36	56
$X \sim LN(0,0.25) + 0.4$, $Y \sim LN(0,1)$	32	43	35	41	35	54	60	59	37	65	85	85

Table D.6: Estimated power (%) of the van der Waerden (vdW), WMW (W), FP, and the combined (C) tests. LN: lognormal distribution.

Figure D.3 shows the density functions of $Gamma\,(5, rate = 1)$ and $Gamma\,(2, rate = 0.5)$. Table D.7 shows the result of the power study from Section 4 using these two distributions.

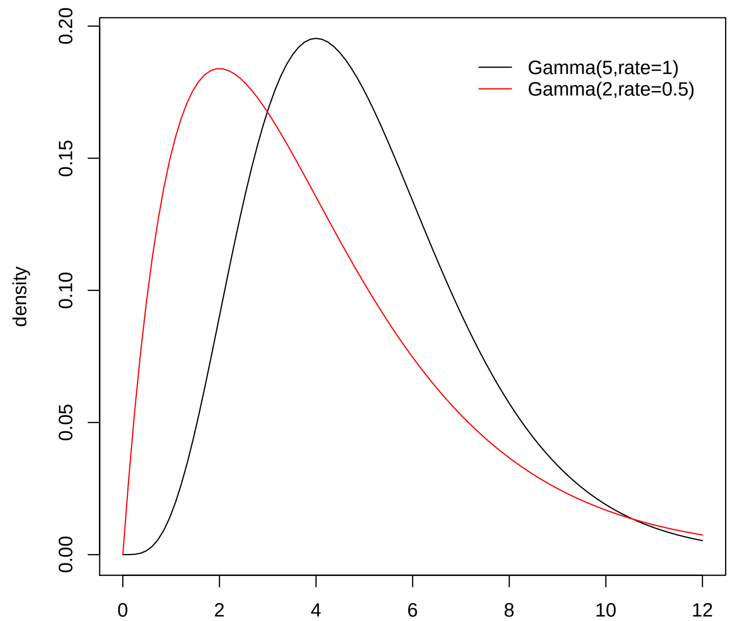

Figure D.3: Densities of two Gamma distributions.

	$n = m$				$n = 2m$				$n = 4m$			
	vdW	W	FP	C	vdW	W	FP	C	vdW	W	FP	C
normal approx., $m = 30$												
$X \sim G(5,1)$, $Y \sim G(2,0.5)$	50	51	52	52	66	63	59	63	75	70	63	70
$X \sim G(2,0.5)$, $Y \sim G(5,1)$	52	52	53	54	61	65	71	71	69	76	84	84
permutation, $m = 30$												
$X \sim G(5,1)$, $Y \sim G(2,0.5)$	51	52	51	51	66	63	58	62	75	70	61	67
$X \sim G(2,0.5)$, $Y \sim G(5,1)$	52	53	52	52	62	65	70	69	69	76	82	82

Table D.7: Estimated power (%) of the van der Waerden (vdW), WMW (W), FP, and the combined (C) tests. G: gamma distribution.

D.3 Additional Monte Carlo output

Table D.8 and D.9 repeat the simulation studies of Section 2 under the scenario that the distributions of X and Y have different means.

m	$n = m$			$n = 2m$			$n = 4m$		
	FP*	(2)	(3)	FP*	(2)	(3)	FP*	(2)	(3)
$s_y = 1$									
5	11.9	11.9	10.2	7.7	10.7	8.0	-2.5	6.0	4.0
10	7.4	7.4	4.4	2.0	3.5	1.2	-3.0	1.3	-0.3
20	1.5	1.5	-0.5	0.3	1.1	-0.4	-2.7	-0.6	-1.5
30	3.1	3.1	1.6	0.8	1.3	0.3	-1.9	-0.4	-1.0
50	0.4	0.4	-0.5	0.6	1.0	0.3	0.3	1.2	0.8
$s_y = 2$									
5	8.5	8.5	8.0	14.5	11.1	8.5	7.2	7.0	4.3
10	5.8	5.8	3.6	5.4	3.7	1.4	2.7	2.6	0.7
20	1.8	1.8	0.3	4.0	3.1	1.7	0.1	0.0	-1.1
30	3.5	3.5	2.4	1.9	1.4	0.4	-0.3	-0.3	-1.1
50	-0.2	-0.2	-1.0	0.5	0.2	-0.5	1.8	1.8	1.3
$s_y = 0.5$									
5	7.5	7.5	6.6	-1.5	6.0	4.8	-10.5	2.4	1.5
10	5.3	5.3	3.0	-1.4	2.4	0.9	-6.1	0.4	-0.5
20	0.1	0.1	-1.5	-2.0	-0.1	-1.1	-4.4	-1.2	-1.7
30	2.0	2.0	0.7	-0.8	0.5	-0.2	-2.9	-0.7	-1.1
50	1.5	1.5	0.7	0.4	1.2	0.8	-0.4	0.9	0.7

Table D.8: Percent bias of three variance estimators. FP*: the original variance estimator from Fligner and Policello (1981); (2) and (3): variance estimators based on formula (2) and (3). $X \sim Logistic(0, 1)$, $Y \sim Logistic(1, s_y)$.

m	$n = m$			$n = 2m$			$n = 4m$		
	FP*	(2)	(3)	FP*	(2)	(3)	FP*	(2)	(3)
$\sigma_y = 1$									
5	14.7	14.7	11.9	10.4	13.5	9.9	-0.1	8.4	5.7
10	10.6	10.6	6.8	6.5	8.1	5.1	0.8	5.2	3.3
20	7.8	7.8	5.2	5.3	6.2	4.4	3.5	5.8	4.7
30	4.4	4.4	2.6	5.0	5.6	4.3	2.2	3.7	2.9
50	4.3	4.3	3.1	2.9	3.3	2.5	1.3	2.2	1.7
$\sigma_y = 2$									
5	9.5	9.5	8.8	15.0	11.0	8.2	10.6	9.1	6.2
10	6.3	6.3	4.1	9.3	7.2	4.7	5.9	5.2	3.1
20	4.4	4.4	2.7	5.5	4.5	3.0	5.6	5.3	4.0
30	2.1	2.1	0.9	5.5	4.8	3.7	4.0	3.8	2.9
50	3.1	3.1	2.3	3.6	3.2	2.5	2.2	2.0	1.5
$\sigma_y = 0.5$									
5	10.8	10.8	9.0	-0.4	7.4	5.6	-9.7	3.5	2.3
10	9.2	9.2	6.2	1.7	5.8	3.9	-3.2	3.6	2.6
20	6.5	6.5	4.4	2.6	4.7	3.5	0.3	3.8	3.1
30	4.7	4.7	3.2	2.8	4.2	3.4	0.2	2.5	2.0
50	3.8	3.8	2.8	1.2	2.0	1.5	0.0	1.3	1.1

Table D.9: Percent bias of three variance estimators. FP*: the original variance estimator from Fligner and Policello (1981); (2) and (3): variance estimators based on formula (2) and (3). $X \sim N(0,1)$, $Y \sim N(1, \sigma_y^2)$.

Figure D.4 is referred to in Section 4 to demonstrate the biasedness of the variance estimator $\min\left(\hat{V}, V_{mn}\right)$.

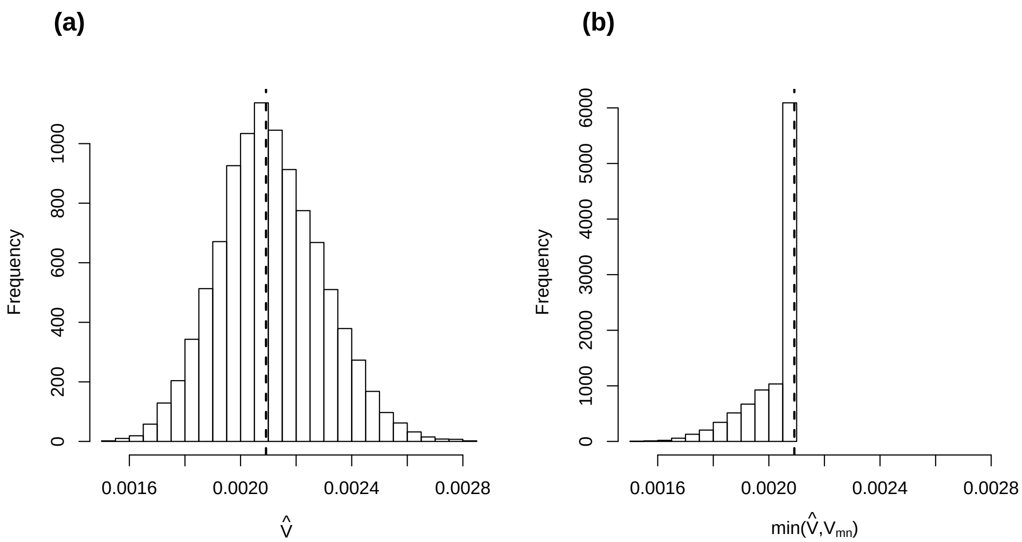

Figure D.4: Sampling distributions of $\hat{V}$, the variance estimator based on formula (2), and $min(\hat{V}, V_{mn})$. $X, Y \sim N(0, 1)$. The vertical line indicates the theoretical variance of U, V_{mn}.

E Sources of datasets

The MTCT dataset (*dat.mtct.rob*) is available as part of the robustrank R package that is hosted at the Comprehensive R Archive Network. The air quality dataset (*airquality*) is available as part of the standard R distribution.

References

Fligner, M.A. and Policello, G.E. (1981), “Robust rank procedures for the Behrens-Fisher problem,” *Journal of the American Statistical Association*, 76, 162–168.

Hollander, M., Wolfe, D. and Chicken, E. (2013), *Nonparametric Statistical Methods*, Wiley Series in Probability and Statistics, Wiley.

Lee, A. (1990), *U-statistics: Theory and Practice*, CRC Press, Boca Roton, FL.