INSTITUTE for REPLICATION

No. 55
I4R DISCUSSION PAPER SERIES

Replication Report: A Comment on Peter T. Leeson, R. August Hardy and Paola A. Suarez (2022)

Lucas Braga de Melo

Giulio G. Cantone

Jacopo Gabani

Xiaoying Gao

August 2023

INSTITUTE for REPLICATION

I4R Discussion Paper Series

I4R DP No. 55

Replication Report: A Comment on Peter T. Leeson, R. August Hardy and Paola A. Suarez (2022)

Lucas Braga de Melo[1], Giulio G. Cantone[2], Jacopo Gabani[3], Xiaoying Gao[3]

[1]Nottingham Interdisciplinary Centre for Economic and Political Research, University of Nottingham/United Kingdom
[2]University of Sussex Business School, Falmer, Brighton/United Kingdom
[3]University of York/United Kingdom

August 2023

ISSN: 2752-1931

Replication Report: A Comment on Peter T. Leeson, R. August Hardy and Paola A. Suarez (2022).

[equal contribution, alphabetical order by last name]

Lucas Braga de Melo[1], Giulio G. Cantone[2], Jacopo Gabani[3,4]*, Xiaoying Gao[3]

Abstract

Peter Leeson, August Hardy and Paola Suarez (2022) test maximizing behaviour of panhandlers at several Metrorail stations in Washington, D.C. Their main findings are that "stations with more panhandling opportunities attract more panhandlers" (the first statement) and that "cross-station differences in hourly panhandling receipts are statistically indistinguishable from zero" (the second statement). We test computational reproducibility and robustness replicability of their results. We can reproduce both statements, in Stata and R. Our robustness replications for the first statement confirm the authors' results in the vast majority of cases (replication was successful in 91% of the cases). Our robustness replications for the second statement might raise doubts on this finding. We run weighted ANOVA tests, we change the bounds in minutes used by authors by 5 minutes in their robustness checks, we run Bartlett's tests of equality of variances of means, and run pair-wise tests of equality of means. In three out of four cases we cannot replicate the results, and the differences (of either means, medians or variances of donations) across Metrorail stations are statistically different from zero. We hypothesize that panhandlers have a general idea about which stations have more passers-by, and will rationally go more often there. However, they are unlikely to have information about smaller variations in the number of passers-by (e.g., variations in passers-by at the same station over time due to non-public events), and therefore might find it difficult to perfectly maximize donations.

[1]Nottingham Interdisciplinary Centre for Economic and Political Research, University of Nottingham, Nottingham, UK

[2]Science Policy Research Unit, University of Sussex Business School, University of Sussex, Brighton, UK

[3]Department of Economics and Related Studies, University of York, York, UK

[4]Centre for Health Economics, University of York, York, UK

*corresponding author

1. Introduction

Peter Leeson, August Hardy and Paola Suarez (Leeson et al., 2022), henceforth LHS, test the implication of maximizing behavior amid competition in the panhandling market at Metrorail Stations in Washington, District of Columbia. Their analysis test maximizing behavior among panhandlers. They visited metro stations in Washington and observed panhandles for ten months in 2016–2017 to collect the main data. Their main findings are that "stations with more panhandling opportunities attract more panhandlers" (the first statement) and that "cross-station differences in hourly panhandling receipts are statistically indistinguishable from zero" (the second statement). These results mean that the returns of panhandling are statistically equal across station, which is a fundamental aspect of competitive markets.

In this replication, we test the computational reproducibility (i.e., same data and same procedure as LHS) and perform robustness replications (i.e., same data, and different procedures) of the main LHS results. Firstly, we test the reproducibility of the results using the replication package provided by the authors. Second, we test the replicability of the first and second statements through several robustness tests noted below.

2. Reproducibility, including using a different software

2.1 Reproduction with Stata

We have successfully reproduced the results of LHS in Stata using the package provided. All Tables and Figures could be easily reproduced, and their contents are the same as in the paper. The code script is easy to understand. Tables are not saved in text files in the original script. Despite that, replicating LHS in STATA is straightforward and easy.

However, we note two points. First, in Table 4, the analysis of variance (ANOVA) analysis is shown. LHS swap the rows' names: the "between" metro station sum of squares should be "within" metro stations, and vice versa. Second, several robustness checks are mentioned in Note 16 (i.e., using mean instead of median to test equality of donations and dollars received) and Note 17 (using lower/upper bounds of the length of the observations in minutes), but none of them are present in the replication package.

2.1 Reproduction with R

We also reproduce the entire analysis, from data preparation to analysis, using R. The results of LHS are entirely confirmed. For reproduction in R, we use the six datasets provided by LHS in .csv. These are merged into a singular one and variables are codified following LHS operative definition. Code in Stata was not directly used for the reproduction in R.

Additional notes:

- Of the four independent variables, one is the main regressor and it is a discrete variable, one is an estimator of the friendliness of passers-by, and the other two are binary variables. The two binary variables were originally discrete time variables, then dichotomized.
- The main regression also includes time fixed effects. Errors are clustered at the metro station level, which is not used as a fixed effect. All the independent variables except passers-by are invariant at the metro station level.

We reproduce the results using the package fixest.

Table 1. Reproduction in R

	Original LHS	R reproduction using fixest
Passerby	0.005*** (0.001)	0.005*** (0.001)
Friendliness	1.595 (0.851)	1.595 (0.851)
Shuttle	0.781*** (0.182)	0.781*** (0.182)
Service	0.450 (0.278)	0.450 (0.278)
Time FE (day)	YES	YES
Adj. R2	0.39	0.39

Notes: dependent variable is always number of panhandlers, robust standard errors in parentheses, clustered at the metro station level. *** p<0.01, ** p<0.05, * p<0.1

3. Robustness replication to the first statement

3.1. Monthly or daily measures

We first start by writing the equation used in LHS to justify the statement that panhandlers go where there are more opportunities.

$$Y_{ds} = Passers_{ms} + Friendly_s + Service_s + Shuttle_s + T_d + e_{ds} \quad [1]$$

Where d represents day, s represent station, m represent month, and T_d are day fixed effects.

We note two things from this equation: first, that the dependent variable is measured by day, and the time-variant independent variable (passers-by) is measured by month. Second, we also note that there are no station fixed effects. Station fixed effects would not allow to measure any effect of station time-invariant characteristics. However, they would allow to account for any time-invariant differences across the stations. LHS do mention that a number of characteristics of metro stations are similar or are dependent on "average" passers-by (e.g., number of garbage bins).

In the first instance, we simply reproduce the same equation, measuring the dependent variable at the month level, as shown in Eq. [2].

$$Y_{ms} = Passers_{ms} + Friendly_s + Service_s + Shuttle_s + T_m + e_{ms} \qquad [2]$$

We would expect Eq. [2] to provide very similar results (using Y_{ds} should only improve precision), which it does. We note that, when using Eq. [2], results are very similar if we use day fixed effects T_d instead of month fixed effect T_m.

3.2 Metro station and season fixed effects

Secondly, we add metro station fixed effects S_s, as shown in Eq. [3].

$$Y_{ms} = Passers_{ms} + S_s + T_d + e_{ms} \qquad [3]$$

This would allow to see whether panhandlers respond to differences in passers-by over time within each of the 26 stations in the sample.

There are issues related to this robustness check. First, we cannot include in the equation time-invariant variables such as distance to homeless services, distance to homeless shuttle stops, and friendliness. Second, in the full sample, the within-variance of passers-by is substantially less than the between-variance across stations.

However, there are also reasons that might justify this choice. The within-station variance in passers-by is about a fifth of the between-stations passers-by variance. This suggests that over-time variation in passers-by is limited, but not completely marginal. There might also be time-invariant differences in the stations architecture, for example driven by the year of construction or renovation, that are not visible to economists, but might be visible to built-environment experts. Finally, we note that LHS runs the Kruskall-Wallis test of equality of median across stations. The test is used to measure equality of medians across different *populations*: if panhandlers are assumed to be from a different population depending on the station they visit, then it seems coherent to include station fixed effects.

We present the results in Table 2. Column 1 provides the original point estimates from LHS. We do not find that the number of panhandlers is associated with the number of passers-by, when including metro station fixed effects. This does not confute the finding that panhandlers act rationally (i.e., more panhandlers go to stations where *usually* there are more passers-by). However, it qualifies the statement: if the number of passers-by at a given station increases or decreases over time in a way that is not clearly visible to panhandlers, then the number of panhandlers does not vary according to the number of passers-by. We do not find this behaviour to be irrational, but simply the result of limited information: panhandlers do know that some stations have *usually* more passers-by, but they are not able to know and/or predict small monthly changes in the number of passers-by at any given station.

In addition to metro fixed effects, we also include season fixed effects in column 4 to check whether the season might affect LHS results. Using season fixed effects confirms the original LHS results.

Table 2. Season fixed effects, metro fixed effects and monthly independent variable

	(1) Original LHS	(2) Monthly measures	(3) Metro FE	(4) Season FE
Number of passers-by	0.005***	0.004***	-0.002	0.004***
	(0.001)	(0.001)	(0.006)	(0.001)
Passers-by friendliness	1.595*	1.391*	2.582	1.381
	(0.851)	(0.786)	(26.756)	(0.853)
Homeless shuttle stop	0.781***	0.673***	0.942	0.654***
	(0.182)	(0.173)	(6.766)	(0.177)
Homeless service provider	0.450	0.437	0.644	0.422
	(0.278)	(0.262)	(4.654)	(0.292)
Observations	242	242	242	239
Adjusted R-squared	0.335	0.599	0.442	0.355
Season FE	NO	NO	NO	YES
Metro station FE	NO	NO	YES	NO

Notes: dependent variable is number of panhandlers, robust standard errors in parentheses, clustered at the metro station level. *** p<0.01, ** p<0.05, * p<0.1

3.3 Changes to variables used

The table below (Table 3) includes the replication of the main regression, removing variables backwards step by step. The analysis could observe the effect of time variables and the main influencing determinants. All the results were conducted by OLS regression with SEs clustered by metro station and include day fixed effects, except for the Pooled OLS model. The replication results show that the original model results are robust to all the checks conducted.

Table 3. Changing variables included in the model

	(1)	(2)	(3)	(4)	(5)
	Original LHS	Pooled OLS	Model 1	Model 2	Model 3
Number of passersby	0.005*** (0.001)	0.004*** (0.001)	0.005*** (0.001)	0.005*** (0.001)	0.005*** (0.001)
Passerby friendliness	1.595 (0.851)	1.346 (0.841)	1.113 (0.790)		
Homeless shuttle stop	0.781*** (0.182)	0.693*** (0.175)	0.820*** (0.243)	0.711*** (0.213)	
Homeless service provider	0.450 (0.278)	0.438 (0.290)			
Time (day) FE	YES	NO	YES	YES	YES
R squared	0.39	0.33	0.37	0.36	0.32
Obs	242	242	242	242	242

Notes: ***p<0.01; **p<0.05, *p<0.1, robust standard errors in parenthesis clustered at metro station level, dependent variable is number of panhandlers in all models

3.4 Using alternative regression forms: negative binomial, interactive models

As an alternative regression form, Negative Binomial (NB) is tested instead of OLS.

The shuttle and service variables are dichotomized in LHS using an arbitrary threshold of 10 minutes (shuttle = 1 if a shuttle stop is within 10 minutes, 0 otherwise, same for homeless services). However, they are counts of minutes. Therefore, we run models using these two variables in their original count form. In both these cases, the original LHS results are robust to the changes.

Table 4 Models that take into consideration the discrete nature of the outcome variable

	(1)	(2)	(3)
	Original LHS	Negative binomial	Negative binomial
Number of passer-by	0.005***	0.005***	0.004***
	(0.001)	(0.001)	(0.001)
Friendliness of passersby	1.595	1.445	1.379
	(0.851)	(0.902)	(1.123)
Shuttle stop – dichotomized	0.781*** (0.182)	0.615* (0.250)	
Service provider – Dichotomized	0.450 (0.278)	0.280 (0.311)	
Shuttle stop – discrete minutes			0.544 (0.315)
Service provider – discrete minutes			0.294 (0.285)
Adj. R2	0.39	0.35	0.05
Obs	242	242	242

Notes: dependent variable is number of panhandlers, robust standard errors in parentheses, clustered at the metro station level. *** p<0.01, ** p<0.05, * p<0.1

Besides binomial models, we have checked the robustness of statement 1 using interaction terms and adding a square-root version of the main independent variable (number of passers-by).

In Table 5, the non-linear model contains the square root of the number of passers-by, and the interactive models add an interaction term between the presence of services / shuttle stop nearby, and the number of passers-by.

The equations used for these robustness replications are listed below:

$$Y_{ds} = Passers_{ms} + \sqrt{Passers_{ms}} + Friendly_s + Service_s + Shuttle_s + T_d + e_{ds} \qquad [4]$$

$$Y_{ds} = Passers_{ms} + Friendly_s + Service_s + Shuttle_s + Passers_{ms} * Shuttle_s + T_d + e_{ds} \qquad [5]$$

$$Y_{ds} = Passers_{ms} + Friendly_s + Service_s + Shuttle_s + Passers_{ms} * Service_s + T_d + e_{ds} \qquad [6]$$

Table 5. Comparison among different formulations of models

	(1)	(2)	(2)	(3)
Dependent variable: Number of panhandlers	Original LHS	Non-linear model	Interactive model	Interactive model
Number of passers-by	0.005***	0.008*	0.004***	0.004***
	(0.001)	(0.004)	(0.001)	(0.001)
Friendliness of passers-by	1.595	1.474	1.524	1.535
	(0.885)	(1.013)	(0.998)	(0.873)
Shuttle stop	0.781***	0.739***	0.466	0.819***
	(0.173)	(0.152)	(0.518)	(0.133)
Homeless service	0.450	0.442	0.452	-0.344
	(0.264)	(0.270)	(0.264)	(0.395)
Square root of passers-by		-0.108		
		(0.142)		
Passersby * shuttle			0.001	
			(0.002)	
Passersby * service				0.002**
				(0.001)
Adjusted R2	0.39	0.39	0.39	0.40
N	242	242	242	242
Date fixed effects	YES	YES	YES	YES

Notes: Observations are Metro station-visits. Robust standard errors clustered by Metro station in parentheses except for (2). $p<0.1$*, $p<0.05$**, $p<0.01$***

The results confirm the first statement of the paper. The coefficient for passers-by in Column 2 is not significant: this suggests that the linear model with interactive terms or with time fixed effects are the best choice to explain the relationship between panhandlers and passers-by. In model (4), the number of passersby, the presence of a shuttle stop nearby, and the interactive item of the number of passersby and the presence of homeless services nearby are significant at 0.01 level. We find that this is a potentially important result: panhandlers are more responsive to the number of passers-by when there are homeless services nearby.

Overall, our robustness replications for the first statement confirm the authors' results in the vast majority of cases (replication was successful in 91% of the cases).

4. Robustness replications for the second statement

4.1. Test of equality of means: pair-wise tests

One of the main results of LHS is that returns of panhandles are equal across metro station. If the mean of donation in dollars are equal across metro stations, then panhandlers are maximizing donations by choosing the most profitable metro station. Therefore, it is important to conduct a robustness test on the evidence of means equality reported in LHS.

Table 6 consists of a robustness check of the ANOVA analysis done by LHS. As can be seen in Panel A, LHS don't reject the null hypothesis that all metro stations have equal means. Panel B do pair-wise Wald tests For example, the first test has a joint null hypothesis that stations Farragut North and Farragut have the same mean joint with the null hypothesis that stations West Farragut and West Gallery Place also have the same mean. The alternative hypothesis is that at least one of these pairs of stations have different means. Results show evidence that none of the pairwise selected stations have different means.

Table 6. Pair-wise test of equality of means

Analysis of variance ANOVA hourly donations in dollars (original LHS)

	Panel A - Table 4 of LHS		
		F	Prob > F
		0.400	0.809
Source	SS	df	MS
Between groups	16.569	4	4.142
Within groups	779.724	75	10.396
Total	796.293	79	10.080

Pair-wise means test equality hourly donations in dollars

	Panel B – Robustness test		
1a) H0:	Farragut North - Farragut West = 0	Chi2	Prob>Chi2
1b) H0:	Farragut West - Gallery Place = 0		
H1:	At least two stations have different means	0.760	0.685
2a) H0:	Gallery Place - McPherson Square = 0		
2b) H0:	McPherson Square - Metro Center = 0		
H1:	At least two stations have different means	0.680	0.712

Note: Table 6 Panel B consists of a robustness check of the ANOVA analysis done by LHS. As can be seen in Panel A, LHS cannot reject the null hypothesis that all metro stations have equal means. Panel B do pair-wise Wald tests. For example, the first test has a joint null hypothesis that Farragut North and Farragut have the same mean joint with the null hypothesis that West Farragut and West Gallery Place have the same mean. The alternative hypothesis is that at least two stations have different means. Results show evidence that none of the select pairwise mean is different from each other.

4.2. Test of equality of means via ANOVA: hourly and by minute dollars and donations

The authors compute the hourly donations and dollars by dividing by 60 the number of donations and dollars donated (be them in cash or in kind) received per minute (for example, if an observation lasted for four minutes, and 1 dollar is received, then that observation is recorded as "15 dollars per hour"). This however results in an implicit choice: each observation is given equal weight, regardless of the minutes it lasted. In other words, a four-minutes observation has the same importance as a 180-minutes (or any number of minutes) observation. However, one might argue that observations that last longer should have more weight. We can show why in a simple example. Presume we visit each of the five stations selected by LHS once. In one case, we visit the station for 4-minutes, and we find a 1 US$ donation: according to LHS methodology, this station shows 15 US$ per hour. We visit the other four stations once, for 60 minutes, and we find no donations at all: according to LHS methodology, these stations show 0 US$ per hour. The average hourly donation across the five stations, according to LHS (i.e., without weighting), would be 3 US$, even if panhandlers actually received 1 US$ in more than four hours. If we weight by the length of each observation in minutes, we have an average donation of 0.25 US$, which reflects the actual amount received per waiting time.

Given the above points, we believe that it is relevant to check whether using weights equivalent to the length of each observation in minutes affect the results. Using both weighted and unweighted analyses is also recommended in the literature (Solon et al., 2015). The results of this check are below. The hypothesis that hourly dollars and hourly donations are equal is rejected (in both cases, $p<0.01$).

Table 7. ANOVA tests of equality of means of hourly donations and hourly dollars received, using weights

PANEL A: hourly donations	Sum of Squares	Degrees of freedom	Mean square	ANOVA F statistics (weighted)	ANOVA F statistics (Original study)
Between groups	4806	4	1201	192***	0.40
Within groups	25383	4050	6	(0.000)	(0.809)

	Sum of Squares	Degrees of freedom	Mean square	ANOVA F statistics (weighted)	ANOVA F statistics (Original study)
Total	30190	4054	7		
PANEL B: hourly dollars					
	Sum of Squares	Degrees of freedom	Mean square	ANOVA F statistics (weighted)	ANOVA F statistics (Original study)
Between groups	25806	4	6451	94.72***	0.15
Within groups	224298	3293	68	(0.000)	(0.962)
Total	250104	3297	76		

Notes: One-way ANOVA test of equality of means with *p*-values in parentheses; p<0.01***

4.3. Tests of equality of variances across stations: mean, median, bounds and Bartlett's test

LHS test equality of mean donations and dollars, equality of median donation and dollars, equality of variance of median donations and dollars (in all cases, across stations), and subject all these results to several robustness checks.

First, we note that these tests are not present in the replication package provided therefore we have tested them based on what we understood from the paper. All the statements made by LHS are reproduced.

Second, because LHS only test for the equality of the variance of mean hourly donations and dollars once, for the full sample, and do not mention any further robustness checks, we conduct further tests for the equality of the variance of mean donations and dollars. This seems only normal for three reasons. First, because LHS use the equality of medians *and* means to test their second statement. Second, because LHS also test the equality of the variance of means dollars and donations, but they do not check the robustness of that result in the same way they do for their other results. Third, conceptually, while using the median fits better a skewed distribution, it might be less valid because it is not a "median" donation that panhandlers put in their pockets: total income of panhandlers is equivalent to the mean receipts' times number of donations.

Regarding equality of variance of mean donations, and median donations, changing the lower bound from 10 to 15 minutes result in the hypothesis of equality of variances of mean donations and median donations to be rejected (see Panel A, Table 8)

Regarding equality of variances of mean dollars, using the Bartlett's test of homogeneous variance does not confirm the statement that the variances of mean receipts in US$ are homogeneous across stations. This test is actually produced by the replication package provided by LHS, therefore we are only reporting a test produced by the author's themselves (see Panel B, Table 8)

Table 8. Robustness checks for equality of variance of mean, and equality of variance of median

PANEL A: HOURLY DONATIONS

		Remove observations lasting below or equal to;	
	Full sample (LHS)	10min	15min
W-statistics: test of equality of variance of mean	1.66	2.63**	3.41**
W-statistics: test of equality of variance of median	0.54	0.95	2.66**
Bartlett's test: equality of variance of the mean	4.58		
Observations	80	73	64

PANEL B: HOURLY DOLLARS

		Remove observations lasting below or equal to;	
	Full sample (LHS)	10min	15min
W-statistics: test of equality of variance of mean	1.16	1.63	1.96
W-statistics: test of equality of variance of median	0.28	0,49	0.73
Bartlett's test: equality of variance of the mean	16.3***		
Observations	67	61	54

5. Conclusion

All the results mentioned by LHS in the paper are reproduced successfully in Stata and R. However, there is a mistake in rows' naming in Table 3.

Our analysis confirms the first statement made by LHS that "panhandlers go where there are more opportunities" in all eleven robustness replications, except one (replication was successful in 91% of the cases). Only when using metro station fixed effects LHS results are not confirmed. However, we note that using fixed effects in this case is problematic given the importance of time-invariant variables and there are interactive effects between passers-by and homeless services.

We also explore the robustness of the second statement that “cross-station differences in hourly panhandling receipts are statistically indistinguishable from zero”. For this statement, one robustness replication (out of four) is successful. This may cast some doubt on the robustness of this statement. LHS gave equal-weight to each observation, regardless of their length in minutes (i.e., the regression is unweighted). Following the literature (Solon et al., 2015), we run both weighted and unweighted regressions, and LHS results are not confirmed in the weighted ANOVA tests. Secondly, we repeated the robustness checks done by LHS (i.e., removing observations shorter than X minutes when measuring equality of variances) on the median donations and median dollar received, using means instead of median values: also in this case, we found that the main result was not robust to all checks. Thirdly, the Bartlett’s test, a standard test of the equality of the variance of the mean, does not confirm LHS results. However, the pair-wise test of equality of mean dollars and donations does confirm original LHS results.

We hypothesize that panhandlers have a general idea about which stations have more passers-by. However, they are unlikely to have information about smaller variations in the number of passers-by (e.g., variations in passers-by at the same station over time due to not entirely public events), and therefore might find it difficult to perfectly maximize donations. This hypothesis seems to be supported by the specification using metro station fixed effects.

References

Leeson, P. T., Hardy, R. A., & Suarez, P. A. (2022). Hobo Economicus*. *The Economic Journal*, *132*(646), 2325–2338. https://doi.org/10.1093/EJ/UEAB103

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human Resources*. https://doi.org/10.3368/jhr.50.2.301