The importance of reproducibility in COVID-19 research: the case of population density and the spread of the pandemic

Antonio Paez [1] [*]

1 School of Earth, Environment and Society, 1280 Main St West, Hamilton, Ontario L8S 4K1 Canada

* Corresponding author: paezha@mcmaster.ca

Abstract

The emergence of the novel SARS-CoV-2 coronavirus and the global COVID-19 pandemic has led to explosive growth in scientific research. Given the high stakes of the situation, it is essential that scientific activites, on which good policy depends, are as transparent and reproducible as possible. Reproducibility is key for the efficient operation of the self-correction mechanisms of science, which work to weed out errors and refine our understanding of social and physical phenomena. In this paper, the importance of reproducibility is illustrated for the case of the association between population density and the the spread of SARS-CoV-2. Transparency and openness means that the same problem can, with relatively modest efforts, be examined by independent researchers who can verify findings, and bring to bear different perspectives, approaches, and methods—sometimes with consequantial changes in the conclusions, as the empirical example in this paper shows.

Introduction

The emergence of the novel SARS-CoV-2 coronavirus in 2019, and the global pandemic that followed in its wake, led to an explosive growth of research. According to Fraser et al. [1], over 125,000 COVID-19-related papers were released in the first ten months from the first confirmed case of the disease. Of these, more than 30,000 were shared in pre-print servers, the use of which also exploded in the past year [2–4].

Given the heavy human and economic cost of the pandemic, there has been a natural tension in the scientific community between the need to publish research results quickly and the imperative to maintain consistently high quality standards in scientific reporting; indeed, a call for maintaining the standards in published research has even called this deluge of publications a “carnage of substandard research” [5]. Part of the challenge of maintaining quality standards in published research is that, despite an abundance of recommendations and guidelines [6–9], in practice reproducibility has remained a lofty but somewhat aspirational goal [10,11]. As reported in the literature, only a woefully small proportion of published research was actually reproducible before the pandemic [12,13], a situation that does not appear to have changed substantially since [14,15].

The push for open data and software, along with more strenuous efforts towards open, reproducible research, is simply a continuation of long-standing scientific practices of independent verification. Despite the (at times disproportionate) attention that high profile scandals in science tend to elicit in the media, science as a collective endeavor is remarkable for being a self-correcting enterprise, one with built-in mechanisms and incentives to weed out erroneous ideas. Over the long term, facts tend to prevail in science. At stake is the shorter-term impacts that research may have in other spheres of economic and social life. The case of economists Reinhart and Rogoff comes to mind: by the time the inaccuracies in their research were uncovered [see 16], their claims about debt and economic growth had already been seized by policy-makers on both sides of the Atlantic to justify austerity policies in the aftermath of the Great Recession of 2007-2009[1]. As later research has demonstrated, those policies cast a long shadow, and their sequels continued to be felt for years [17].

In the context of COVID-19, a topic that has grabbed the imagination of numerous thinkers has been the prospect of life in cities after the pandemic [18]. The fact that the worst of the pandemic was initially felt in dense population centers such as Wuhan, Milan, Madrid, and New York, prompted a flurry of research into the associations between density and the spread of the pandemic. Some important questions hang on the results of these research efforts. For example, are lower density regions safer from the pandemic? Are de-densification policies warranted, at least in the short term? And in the longer term, will the risks of life in high density regions presage a flight from cities? Over the past year, numerous papers have sought to throw light into the underlying issue of density and the pandemic; nonetheless the results, as will be detailed next, remain mixed. Further, to complicate matters, precious few of these studies appear to be sufficiently open to support independent verification.

The objective of this paper is to illustrate the importance of reproducibility in research, particularly in the context of the flood of COVID-19 papers. To this end, a recent study by Sy et al. [19] is chosen as an example of reproducible research. The objective is not to malign the analysis of these researchers, but rather to demonstrate the value of openness to allow for independent verification and further analysis. Open data and open code mean that an independent researcher can, with only modest efforts, not only verify the findings reported, but also examine the same data from a perspective which may not have been available to the original researchers due to differences in disciplinary perspectives, methodological traditions, and/or training, among other possible factors. The example, which shows consequential changes in the conclusions reached by different analyses, should serve as a call to researchers to redouble their efforts to increase transparency and reproducibility in research. This paper, in addition, aims to show how data can be packaged in well-documented, shareable units, and code can be embedded into self-contained documents suitable for review and independent verification. The source for this paper is an R Markdown document which, along with the data package, is available in a public repository[2].

Background

The concern with population density and the spread of the virus during the COVID-19 pandemic was fueled, at least in part, by dramatic scenes seen in real-time around the world from large urban centers such as Wuhan, Milan, Madrid, and New York. In theory, there are good reasons to believe that higher density may have a positive association with the transmission of a contagious virus. It has long been known that the potential for inter-personal contact is greater in regions with higher density [see for example the research on urban fields and time-geography 20,21,22]. Mathematically, models of exposure and contagion indicate that higher densities can catalyze the

[1] Nobel Prize in Economics Paul Krugman noted that "Reinhart–Rogoff may have had more immediate influence on public debate than any previous paper in the history of economics" https://www.nybooks.com/articles/2013/06/06/how-case-austerity-has-crumbled/?pagination=false

[2] https://github.com/paezha/Reproductive-Rate-and-Density-US-Reanalyzed

transmission of contagious diseases [23,24]. Models such as these were likely at the root of messages, by some figures in positions of authority, that low density regions faced lower risks from the pandemic[3].

As Rocklöv and Sjödin [23] note, however, mathematical models of contagion are valid at small-to-medium spatial scales (and presumably, small temporal scales too, such as time spent in restaurants, concert halls, cruises), and the results do not necessarily transfer to larger spatial units and different time scales. There are good reasons for this: while in a restaurant, one can hardly avoid being in proximity to other customers-however, a person can choose to (or be forced to as a matter of policy) not go to a restaurant in the first place. Nonetheless, the idea that high density correlates with high transmission is so intuitive that it is often taken for granted even at larger scales [e.g., 25,26]. At larger scales, however, there exists the possibility of behavioral adaptations, which are difficult to capture in the mechanistic framework of differential equations [or can be missing in agent-based models, 27]; these adaptations, in fact, can be a key aspect of disease transmission.

A plausible behavioral adaptation during a pandemic, especially one broadcast as widely and intensely as COVID-19, is risk compensation. Risk compensation is a process whereby people adjust their behavior in response to their *perception* of risk [28–30]. In the case of COVID-19, Chauhan et al. [31] have found that perception of risks in the US varies between rural, suburban, and urban residents, with rural residents in general displaying less concerns about the virus. It is possible that people who listened to the message of leaders saying that they were safe because of low density may not have taken adequate precautions against the virus. People in dense places who could more directly observe the impact of the pandemic may have become overly cautious. Both Paez et al. [32] and Hamidi et al. [33] posit this mechanism (i.e., greater compliance with social distancing in denser regions) to explain the results of their analyses. The evidence available does indeed show that there were important changes in behavior during the pandemic, at least with respect to mobility [34–36]; furthermore, shelter in place orders may have had greater buy-in from the public in higher density regions [37], and the behavior may have persisted beyond the duration of official social-distancing policies [38]. In addition, there is evidence that changes in mobility correlated with the trajectory of the pandemic [39,40]. Given the potential for behavioral adaptation, the question of density becomes more nuanced: it is not just a matter of proximity, but also of human behavior, which is better studied using population-level data and models.

In this respect, the literature to date remains inconclusive.

On the one hand, there are studies that report positive associations between population density and various COVID-19-related outcomes. Bhadra [41], for example, reported a moderate positive correlation between the spread of COVID-19 and population density at the district level in India, however their analysis was bivariate and did not control for other variables, such as income. Similarly, Kadi and Khelfaoui [42] found a positive and significant correlation between number of cases and population density in cities in Algeria in a series of simple regression models (i.e., without other controls). A question in these relatively simple analyses is whether density is not a proxy for other factors. Other studies have included controls, such as Pequeno et al. [43], a team that reported a positive association between density and cumulative counts of confirmed COVID-19 cases in state capitals in Brazil after controlling for covariates, including income, transport connectivity, and economic status. In a similar vein, Fielding-Miller et al. [44] reported a positive relationship between the absolute number of COVID-19 deaths and population density (rate) in rural counties in the US. Roy and

[3]Governor Kristi Noem of South Dakota, for example, claimed that sparse population density allowed her state to face the pandemic without the need for strict policy interventions `https://www.inforum.com/lifestyle/health/5025620-South-Dakota-is-not-New-York-City-Noem-defends-lack-of-statewide-COVID-19-restrictions`

Ghosh [45] used a battery of machine learning techniques to find discriminatory factors, and a positive and significant association between COVID-19 infection and death rates in US states. Wong and Li [46] also found a positive and significant association between population density and number of confirmed COVID-19 cases in US counties, using both univariate and multivariate regressions with spatial effects. Most recently, Sy et al. [19] reported that the basic reproductive number of COVID-19 in US counties tends to increase with population density, but at a decreasing rate at higher densities.

On the gripping hand, a number of studies report non-significant or negative associations between population density and COVID-19 outcomes. This includes the research of Sun et al. [47] who did not find evidence of significant correlation between population density and confirmed number of cases per day *in conditions of lockdown* in China. This finding echoes the results of Paez et al. [32], who in their study of provinces in Spain reported non-significant associations between population density and infection rates in the early days of the first wave of COVID-19, and negative significant associations in the later part of the first lockdown. Similarly, [48] found zero or negative associations between population density and infection numbers/deaths by country. Fielding-Miller et al. [44] found a negative relationship between COVID-19 deaths and population density urban counties in the US. In their investigation of doubling time, White and Hébert-Dufresne [49] identified a negative and significant correlation between population density and doubling time in US states. Likewise, [50] fond a small negative (and significant) association between population density and COVID-19 morbidity in districts in Tehran. And two of the most complete studies in the US by Hamidi et al. [33,51] used an extensive set of controls to find negative and significant correlations between density at the level of counties in the US and COVID-19 cases and fatalities.

As can be seen, these studies are implemented at different scales in different regions of the world. They also use a range of techniques, from correlation analysis, to multivariate regression, spatial regressions, and machine learning techniques. This is natural and to be expected: individual researchers have only limited time and expertise. This is why reproducibility is important. To pick an example (which will be further elaborated in the following sections), the study of Sy et al. [[19]; hereafter SWN] would immediately grab the attention of a researcher with expertise in spatial modelling. Such an expert would likely ask some of the following questions: how were missing counties treated? Is it possible to spatially interpolate missing observations? What are the implications of the spatial sampling framework used in the analysis? Was there evidence of spatial autocorrelation in the residuals of the models? These are questions that in most cases would not occur to a researcher who has not been exposed to spatial statistics or spatial econometrics. Nonetheless, they are relevant and important. Fortunately, SWN give an example of a reasonably open, reproducible research product: their paper is accompanied by (most of) the data and (most of) the code used in the analysis. This means that an independent expert can, with only a moderate investment of time and effort, replicate the results in the paper, as well as ask additional questions.

Alas, reproducibility is not necessarily the norm.

There are various reasons why a project can fail to be reproducible. In some cases, there might be legitimate reasons to withhold the data, perhaps due to confidentiality and privacy reasons [e.g., 52]. But in many other cases the data are publicly available, as has been commonly the case with population-level COVID-19 information. Often the provenance of the data is documented, but the data themselves are not shared [25,33,41,44,51,e.g., 53,54–56]. As any researcher can attest, whether a graduate student or a seasoned scientist, collecting, organizing, and preparing data for a project can take a substantial amount of time. Pointing to the sources of data, even when these sources are public, is a small step towards reproducibility-but only a very small one. Faced with the prospect of having to recreate a data set from raw sources is probably sufficient to dissuade all but the most dedicated (or stubborn) researcher. This is true even if part of the data are shared [e.g., 46]. In other cases, data are shared, but the processes to document the preparation of the data are not fully documented [48,e.g., 57]. These processes matter, as shown by the errors in the spreadsheets of Reinhart and Rogoff [16], and the data of biologist Jonathan Pruitt that led to an "avalanche" of paper retractions[4]. Another situation is when papers share well-documented data, but fail to provide the code used in the analysis [43,e.g., 58,59]. Making code available only "on demand" [e.g., 60] is an unnecessary barrier when most journals offer the facility to share supplemental materials online. Then there are those papers that strive towards reproducibility, sharing well-documented processes and data, as well as the code used in any analyses reported [e.g., 32,37,49,61].

Replicating SWN

Fit (mixed) linear models as in Sy et al.:

```
table1model1 <- lme(R ~ density_log ,
                    random = ~ 1| state,
                    data = county_geo_clean %>%
                      filter(R > 0))
summary(table1model1)
```

```
## Linear mixed-effects model fit by REML
##   Data: county_geo_clean %>% filter(R > 0)
##        AIC      BIC    logLik
##   2383.621 2403.808 -1187.811
##
## Random effects:
##  Formula: ~1 | state
##         (Intercept)  Residual
## StdDev:    0.165504 0.6648325
##
## Fixed effects:  R ~ density_log
##                 Value  Std.Error   DF  t-value p-value
## (Intercept) 2.2740045 0.05468629 1099 41.58272       0
## density_log 0.1621127 0.01485543 1099 10.91269       0
##  Correlation:
##             (Intr)
## density_log 0.797
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.3847932 -0.5972261 -0.1682299  0.3681864 16.2716371
##
## Number of Observations: 1151
## Number of Groups: 51
```

```
intervals(table1model1)
```

```
## Approximate 95% confidence intervals
##
```

[4] https://doi.org/10.1038/d41586-020-00287-y

```
##  Fixed effects:
##                 lower      est.     upper
## (Intercept) 2.1667032 2.2740045 2.3813058
## density_log 0.1329645 0.1621127 0.1912609
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: state
##                     lower     est.     upper
## sd((Intercept)) 0.1080424 0.165504 0.2535261
##
##  Within-group standard error:
##     lower      est.     upper
## 0.6375478 0.6648325 0.6932849
```

Linear mixed models of R0 and density + % private transportation

```
table1model2 <- lme(R ~ density_log + private ,
                   random = ~ 1| state,
                   data = county_geo_clean %>%
                     filter(R > 0))
summary(table1model2)
```

```
## Linear mixed-effects model fit by REML
##   Data: county_geo_clean %>% filter(R > 0)
##        AIC      BIC    logLik
##   2385.478 2410.707 -1187.739
##
## Random effects:
##  Formula: ~1 | state
##         (Intercept)  Residual
## StdDev:   0.1358322 0.6649714
##
## Fixed effects:  R ~ density_log + private
##                 Value Std.Error   DF   t-value p-value
## (Intercept)  3.347070 0.3418910 1098  9.789874  0.0000
## density_log  0.145258 0.0156188 1098  9.300177  0.0000
## private     -0.012547 0.0039434 1098 -3.181760  0.0015
##  Correlation:
##             (Intr) dnsty_
## density_log -0.222
## private     -0.988  0.344
##
## Standardized Within-Group Residuals:
##        Min          Q1         Med          Q3         Max
## -2.7347592 -0.6005758 -0.1610880  0.3876853 16.2104980
##
## Number of Observations: 1151
## Number of Groups: 51
```

```r
intervals(table1model2)
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##                   lower       est.        upper
## (Intercept)  2.67623619  3.3470697  4.017903160
## density_log  0.11461177  0.1452579  0.175904019
## private     -0.02028416 -0.0125468 -0.004809444
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: state
##                      lower      est.    upper
## sd((Intercept)) 0.08059718 0.1358322 0.228921
##
##  Within-group standard error:
##     lower      est.     upper
## 0.6376445 0.6649714 0.6934695
```

Linear mixed models of R0 and density + % private transportation + median income

```r
table1model3 <- lme(R ~ density_log + private + hincome,
                    random = ~ 1| state,
                    data = county_geo_clean %>%
                      filter(R > 0))
summary(table1model3)
```

```
## Linear mixed-effects model fit by REML
##   Data: county_geo_clean %>% filter(R > 0)
##       AIC     BIC   logLik
##   2393.98 2424.25 -1190.99
##
## Random effects:
##  Formula: ~1 | state
##         (Intercept)  Residual
## StdDev:   0.1373995 0.6651151
##
## Fixed effects:  R ~ density_log + private + hincome
##                  Value Std.Error   DF   t-value p-value
## (Intercept)  3.385550 0.3933810 1097  8.606289  0.0000
## density_log  0.146826 0.0171534 1097  8.559585  0.0000
## private     -0.012707 0.0040548 1097 -3.133786  0.0018
## hincome     -0.003256 0.0151105 1097 -0.215493  0.8294
##  Correlation:
##             (Intr) dnsty_ privat
## density_log  0.027
## private     -0.950  0.211
## hincome     -0.493 -0.412  0.228
##
## Standardized Within-Group Residuals:
```

```
##        Min         Q1        Med         Q3        Max
## -2.7162603 -0.5984817 -0.1626793  0.3887931 16.2029477
##
## Number of Observations: 1151
## Number of Groups: 51
```

```
intervals(table1model3)
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##                   lower         est.       upper
## (Intercept)  2.61368596  3.385550103  4.15741424
## density_log  0.11316880  0.146825984  0.18048317
## private     -0.02066315 -0.012707018 -0.00475089
## hincome     -0.03290488 -0.003256195  0.02639249
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: state
##                    lower      est.     upper
## sd((Intercept)) 0.081461 0.1373995 0.2317504
##
##  Within-group standard error:
##     lower      est.     upper
## 0.6377582 0.6651151 0.6936455
```

Linear mixed models of R0 and density + % private transportation + median income + interaction of density and % private transportation

```
table1model4 <- lme(R ~ density_log*private + hincome,
                   random = ~ 1| state,
                   data = county_geo_clean %>%
                     filter(R > 0))
summary(table1model4)
```

```
## Linear mixed-effects model fit by REML
##   Data: county_geo_clean %>% filter(R > 0)
##       AIC      BIC   logLik
##   2407.04 2442.349 -1196.52
##
## Random effects:
##  Formula: ~1 | state
##         (Intercept)  Residual
## StdDev:   0.1378096 0.6653589
##
## Fixed effects:  R ~ density_log * private + hincome
##                         Value Std.Error   DF   t-value p-value
## (Intercept)          3.415859 0.4157861 1096  8.215424  0.0000
## density_log          0.177645 0.1350641 1096  1.315266  0.1887
## private             -0.013123 0.0044568 1096 -2.944502  0.0033
## hincome             -0.002719 0.0153117 1096 -0.177607  0.8591
```

```
## density_log:private -0.000354 0.0015414 1096 -0.229894  0.8182
##  Correlation: 
##                     (Intr) dnsty_ privat hincom
## density_log          0.323                     
## private             -0.952 -0.386              
## hincome             -0.409  0.105  0.139       
## density_log:private -0.323 -0.992  0.414 -0.158
## 
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max 
## -2.7371582 -0.5968733 -0.1638499  0.3883700 16.1980628 
## 
## Number of Observations: 1151
## Number of Groups: 51
```

```
intervals(table1model4)
```

```
## Approximate 95% confidence intervals
## 
##  Fixed effects:
##                            lower          est.        upper
## (Intercept)          2.600032192  3.4158588336  4.231685475
## density_log         -0.087368297  0.1776452292  0.442658756
## private             -0.021867960 -0.0131231030 -0.004378246
## hincome             -0.032762972 -0.0027194668  0.027324038
## density_log:private -0.003378809 -0.0003543603  0.002670088
## attr(,"label")
## [1] "Fixed effects:"
## 
##  Random Effects:
##   Level: state 
##                      lower      est.     upper
## sd((Intercept)) 0.08169505 0.1378096 0.2324679
## 
##  Within-group standard error:
##     lower      est.     upper 
## 0.6379755 0.6653589 0.6939176
```

Some relevant questions

Alas, despite decades' worth of developments in the field of geographical analysis, not all research presented to date has used proper methods for the study of COVID-19. While the answer to the question "do spatial effects really matter in regression analysis" was definitively answered in the positive at least 40 years ago, in practice many researchers continue to ignore the pitfalls of ignoring them. In this paper, I present a reanalysis of the data used by Sy et al. (2021) to study the correlations between the basic reproductive number of COVID-19 and population density in counties. I highlight two related issues: non-systematic sampling in space and spatial autocorrelation. The reanalysis is based on the use of tobit models to account for non-systematic sampling, and spatially autoregressive tobit models to account for spatial autocorrelation in the data generation process. The reanalysis highlights the importance of openness and reproducibility in COVID-19 research. Finally, the results provide a sobering example of the risks of not using appropriate methods in the analysis of geographical data. Furthermore,

Fit tobit version of models

```r
table1model1 <- censReg(R ~ density_log,
                        left = 0,
                    data = county_geo_clean)
summary(table1model1)
```

```
##
## Call:
## censReg(formula = R ~ density_log, left = 0, data = county_geo_clean)
##
## Observations:
##          Total  Left-censored     Uncensored Right-censored
##           3218           2067           1151              0
##
## Coefficients:
##             Estimate Std. error t value Pr(> t)
## (Intercept)  2.33967    0.09232   25.34  <2e-16 ***
## density_log  0.74323    0.02672   27.82  <2e-16 ***
## logSigma     0.56894    0.02396   23.74  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Newton-Raphson maximisation, 7 iterations
## Return code 8: successive function values within relative tolerance limit (r
## Log-likelihood: -3246.5 on 3 Df
```

Linear mixed models of R0 and density + % private transportation

```r
table1model2 <- censReg(R ~ density_log + private,
                    data = county_geo_clean)
summary(table1model2)
```

```
##
## Call:
## censReg(formula = R ~ density_log + private, data = county_geo_clean)
##
## Observations:
##          Total  Left-censored     Uncensored Right-censored
##           3218           2067           1151              0
##
## Coefficients:
##             Estimate Std. error t value Pr(> t)
## (Intercept) 0.825319   0.583014   1.416 0.15689
## density_log 0.754516   0.027546  27.391 < 2e-16 ***
## private     0.017186   0.006548   2.624 0.00868 **
## logSigma    0.570248   0.023977  23.783 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Newton-Raphson maximisation, 7 iterations
## Return code 8: successive function values within relative tolerance limit (r
## Log-likelihood: -3242.96 on 4 Df
```

Linear mixed models of R0 and density + % private transportation + median income

```
table1model3 <- censReg(R ~ density_log + private + hincome,
                    data = county_geo_clean)
summary(table1model3)
```

```
##
## Call:
## censReg(formula = R ~ density_log + private + hincome, data = county_geo_clea
##
## Observations:
##          Total  Left-censored     Uncensored Right-censored
##           3218           2067           1151              0
##
## Coefficients:
##              Estimate Std. error t value  Pr(> t)
## (Intercept) -3.201908   0.680988  -4.702 2.58e-06 ***
## density_log  0.662526   0.026876  24.651  < 2e-16 ***
## private      0.040886   0.006844   5.974 2.32e-09 ***
## hincome      0.297763   0.025726  11.574  < 2e-16 ***
## logSigma     0.531694   0.023920  22.228  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Newton-Raphson maximisation, 7 iterations
## Return code 8: successive function values within relative tolerance limit (r
## Log-likelihood: -3174.96 on 5 Df
```

Tobit models of R0 and density + % private transportation + median income + interaction of density and % private transportation

```
table1model4 <- censReg(R ~ density_log*private + hincome,
                    data = county_geo_clean)
summary(table1model4)
```

```
##
## Call:
## censReg(formula = R ~ density_log * private + hincome, data = county_geo_clea
##
## Observations:
##          Total  Left-censored     Uncensored Right-censored
##           3218           2067           1151              0
##
## Coefficients:
##                      Estimate Std. error t value  Pr(> t)
## (Intercept)         -3.687016   0.733410  -5.027 4.98e-07 ***
## density_log          0.311841   0.174521   1.787   0.0740 .
## private              0.047454   0.007712   6.154 7.58e-10 ***
```

```
## hincome              0.291536   0.025807  11.297  < 2e-16 ***
## density_log:private  0.004097   0.002028   2.020   0.0434 *
## logSigma             0.530095   0.023924  22.158  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Newton-Raphson maximisation, 7 iterations
## Return code 8: successive function values within relative tolerance limit (r
## Log-likelihood: -3173.008 on 6 Df
```

Spatially autoregressive tobit

Fit spatially autoregressive tobit:

```r
# Fit SAR Tobit
fit_sartobit <- sartobit(R ~ density_log + private + hincome_log,
                         B,
                         ndraw = 1000,
                         burn.in = 200,
                         showProgress = TRUE,
                         data = county_geo_clean,
                         computeMarginalEffects = TRUE)
```

```
## Warning in model.matrix.default(mt, mf, contrasts): non-list contrasts argum
## ignored

## Warning: Function SE_classic_setup moved to the spatialreg package

## Registered S3 methods overwritten by 'spatialreg':
##   method                   from
##   residuals.stsls          spdep
##   deviance.stsls           spdep
##   coef.stsls               spdep
##   print.stsls              spdep
##   summary.stsls            spdep
##   print.summary.stsls      spdep
##   residuals.gmsar          spdep
##   deviance.gmsar           spdep
##   coef.gmsar               spdep
##   fitted.gmsar             spdep
##   print.gmsar              spdep
##   summary.gmsar            spdep
##   print.summary.gmsar      spdep
##   print.lagmess            spdep
##   summary.lagmess          spdep
##   print.summary.lagmess    spdep
##   residuals.lagmess        spdep
##   deviance.lagmess         spdep
##   coef.lagmess             spdep
##   fitted.lagmess           spdep
##   logLik.lagmess           spdep
##   fitted.SFResult          spdep
##   print.SFResult           spdep
```

```
##   fitted.ME_res            spdep
##   print.ME_res             spdep
##   print.lagImpact          spdep
##   plot.lagImpact           spdep
##   summary.lagImpact        spdep
##   HPDinterval.lagImpact    spdep
##   print.summary.lagImpact  spdep
##   print.sarlm              spdep
##   summary.sarlm            spdep
##   residuals.sarlm          spdep
##   deviance.sarlm           spdep
##   coef.sarlm               spdep
##   vcov.sarlm               spdep
##   fitted.sarlm             spdep
##   logLik.sarlm             spdep
##   anova.sarlm              spdep
##   predict.sarlm            spdep
##   print.summary.sarlm      spdep
##   print.sarlm.pred         spdep
##   as.data.frame.sarlm.pred spdep
##   residuals.spautolm       spdep
##   deviance.spautolm        spdep
##   coef.spautolm            spdep
##   fitted.spautolm          spdep
##   print.spautolm           spdep
##   summary.spautolm         spdep
##   logLik.spautolm          spdep
##   print.summary.spautolm   spdep
##   print.WXImpact           spdep
##   summary.WXImpact         spdep
##   print.summary.WXImpact   spdep
##   predict.SLX              spdep

##   |
```

```
summary(fit_sartobit)
```

```
## ----MCMC spatial autoregressive Tobit model ----
## Execution time  = 40.052 secs
##
## N draws         =   1000, N omit (burn-in)=    200
## N observations  =   3218, K covariates    =      4
## # censored values =   2067, # observed values =   1151
## Min rho         = -1.000, Max rho         =  1.000
## --------------------------------------------------
##
##              Estimate  Std. Dev   p-level t-value Pr(>|z|)
## (Intercept) -4.399260  0.737342  0.000000  -5.966 2.69e-09 ***
## density_log  0.562608  0.026714  0.000000  21.061  < 2e-16 ***
## private      0.039826  0.006874  0.000000   5.794 7.54e-09 ***
## hincome_log  1.572387  0.138627  0.000000  11.343  < 2e-16 ***
## sige         2.992428  0.173624  0.000000  17.235  < 2e-16 ***
## rho          0.056389  0.004937  0.000000  11.421  < 2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
impacts(fit_sartobit)
```

```
## --------Marginal Effects--------
##
## (a) Direct effects
##             lower_005 posterior_mean upper_095
## density_log   0.21275        0.23102     0.248
## private       0.01180        0.01635     0.021
## hincome_log   0.55351        0.64566     0.744
##
## (b) Indirect effects
##             lower_005 posterior_mean upper_095
## density_log  -0.29273       -0.27252    -0.251
## private      -0.02504       -0.01929    -0.014
## hincome_log  -0.88010       -0.76178    -0.651
##
## (c) Total effects
##             lower_005 posterior_mean upper_095
## density_log -0.044493      -0.041495    -0.038
## private     -0.003850      -0.002942    -0.002
## hincome_log -0.136033      -0.116125    -0.097
```

Conclusion

Words go here.

References

1.
Fraser N, Brierley L, Dey G, Polka JK, Pálfy M, Nanni F, et al. The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. PLOS Biology. 2021;19: e3000959. doi:10.1371/journal.pbio.3000959

2.
Kwon D. How swamped preprint servers are blocking bad coronavirus research. Nature. 2020;581: 130–132.

3.
Vlasschaert C, Topf JM, Hiremath S. Proliferation of papers and preprints during the coronavirus disease 2019 pandemic: Progress or problems with peer review? Advances in Chronic Kidney Disease. 2020;27: 418–426. doi:10.1053/j.ackd.2020.08.003

4.
Añazco D, Nicolalde B, Espinosa I, Camacho J, Mushtaq M, Gimenez J, et al. Publication rate and citation counts for preprints released during the COVID-19 pandemic: The good, the bad and the ugly. PeerJ. 2021;9: e10927. doi:10.7717/peerj.10927

5.
Bramstedt KA. The carnage of substandard research during the COVID-19 pandemic: A call for quality. Journal of Medical Ethics. 2020;46: 803–807. doi:10.1136/medethics-2020-106494

6.
Broggini F, Dellinger J, Fomel S, Liu Y. Reproducible research: Geophysics papers of the future - introduction. Geophysics. 2017;82. doi:10.1190/geo2017-0918-spseintro.1
7.
Ince DC, Hatton L, Graham-Cumming J. The case for open computer programs. Nature. 2012;482: 485–488. doi:10.1038/nature10836
8.
Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014;383: 166–175. doi:10.1016/s0140-6736(13)62227-8
9.
Brunsdon C, Comber A. Opening practice: Supporting reproducibility and critical spatial data science. Journal of Geographical Systems. 2020; doi:10.1007/s10109-020-00334-2
10.
Konkol M, Kray C. In-depth examination of spatiotemporal figures in open reproducible research. Cartography and Geographic Information Science. 2019;46: 412–427. doi:10.1080/15230406.2018.1512421
11.
Konkol M, Kray C, Pfeiffer M. Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study. International Journal of Geographical Information Science. 2019;33: 408–429. doi:10.1080/13658816.2018.1508687
12.
Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JPA. Reproducible research practices and transparency across the biomedical literature. Plos Biology. 2016;14. doi:10.1371/journal.pbio.1002333
13.
Stodden V, Seiler J, Ma ZK. An empirical analysis of journal policy effectiveness for computational reproducibility. Proceedings of the National Academy of Sciences of the United States of America. 2018;115: 2584–2589. doi:10.1073/pnas.1708290115
14.
Sumner J, Haynes L, Nathan S, Hudson-Vitale C, McIntosh LD. Reproducibility and reporting practices in COVID-19 preprint manuscripts. medRxiv. 2020; 2020.03.24.20042796. doi:10.1101/2020.03.24.20042796
15.
Gustot T. Quality and reproducibility during the COVID-19 pandemic. JHEP Rep. 2020;2: 100141. doi:10.1016/j.jhepr.2020.100141
16.
Herndon T, Ash M, Pollin R. Does high public debt consistently stifle economic growth? A critique of reinhart and rogoff. Cambridge Journal of Economics. 2014;38: 257–279. doi:10.1093/cje/bet075
17.
Basu S, Carney MA, Kenworthy NJ. Ten years after the financial crisis: The long reach of austerity and its global impacts on health. Social Science & Medicine. 2017;187: 203–207. doi:10.1016/j.socscimed.2017.06.026
18.
Florida R, Glaeser E, Sharif M, Bedi K, Campanella T, Chee C, et al. How life in our cities will look after the coronavirus pandemic. Foreign Policy. 2020;1. Available: `https://foreignpolicy.com/2020/05/01/future-of-cities-urban-life-after-coronavirus-pandemic/`
19.

Sy KTL, White LF, Nichols BE. Population density and basic reproductive number of COVID-19 across united states counties. PLOS ONE. 2021;16: e0249271. doi:10.1371/journal.pone.0249271
20.
Moore EG, Brown LA. Urban acquaintance fields: An evaluation of a spatial model. Environment and Planning. 1970;2: 443–454. Available: http://www.envplan.com/abstract.cgi?id=a020443
21.
Moore EG. Some spatial properties of urban contact fields. Geographical Analysis. 1970;2: 376–386.
22.
Farber S, Páez A. Running to stay in place: The time-use implications of automobile oriented land-use and travel. Journal of Transport Geography. 2011;19: 782–793. doi:10.1016/j.jtrangeo.2010.09.008
23.
Rocklöv J, Sjödin H. High population densities catalyse the spread of COVID-19. Journal of Travel Medicine. 2020;27. doi:10.1093/jtm/taaa038
24.
Li R, Richmond P, Roehner BM. Effect of population density on epidemics. Physica A: Statistical Mechanics and its Applications. 2018;510: 713–724. doi:10.1016/j.physa.2018.07.025
25.
Cruz CJP, Ganly R, Li Z, Gietel-Basten S. Exploring the young demographic profile of COVID-19 cases in hong kong: Evidence from migration and travel history data. PLOS ONE. 2020;15: e0235306. doi:10.1371/journal.pone.0235306
26.
Micallef S, Piscopo TV, Casha R, Borg D, Vella C, Zammit M-A, et al. The first wave of COVID-19 in malta; a national cross-sectional study. PLOS ONE. 2020;15: e0239389. doi:10.1371/journal.pone.0239389
27.
Gomez J, Prieto J, Leon E, Rodríguez A. INFEKTA—an agent-based model for transmission of infectious diseases: The COVID-19 case in bogotá, colombia. PLOS ONE. 2021;16: e0245787. doi:10.1371/journal.pone.0245787
28.
Noland RB. PERCEIVED RISK AND MODAL CHOICE - RISK COMPENSATION IN TRANSPORTATION SYSTEM. Accident Analysis and Prevention. 1995;27: 503–521. doi:10.1016/0001-4575(94)00087-3
29.
Richens J, Imrie J, Copas A. Condoms and seat belts: The parallels and the lessons. Lancet. 2000;355: 400–403. doi:10.1016/s0140-6736(99)09109-6
30.
Phillips RO, Fyhri A, Sagberg F. Risk compensation and bicycle helmets. Risk Analysis. 2011;31: 1187–1195. doi:10.1111/j.1539-6924.2011.01589.x
31.
Chauhan RS, Capasso da Silva D, Salon D, Shamshiripour A, Rahimi E, Sutradhar U, et al. COVID-19 related attitudes and risk perceptions across urban, rural, and suburban areas in the united states. Findings. Network Design Lab; 2021; doi:10.32866/001c.23714
32.
Paez A, Lopez FA, Menezes T, Cavalcanti R, Pitta MG da R. A spatio-temporal analysis of the environmental correlates of COVID-19 incidence in spain. Geographical Analysis. 2020;n/a. doi:10.1111/gean.12241
33.

Hamidi S, Sabouri S, Ewing R. Does density aggravate the COVID-19 pandemic? Journal of the American Planning Association. 2020;86: 495–509. doi:10.1080/01944363.2020.1777891
34.
Jamal S, Paez A. Changes in trip-making frequency by mode during COVID-19. Findings. Network Design Lab; 2020; doi:10.32866/001c.17977
35.
Harris MA, Branion-Calles M. Changes in commute mode attributed to COVID-19 risk in canadian national survey data. Findings. Network Design Lab; 2021; doi:10.32866/001c.19088
36.
Molloy J, Tchervenkov C, Hintermann B, Axhausen KW. Tracing the sars-CoV-2 impact: The first month in switzerland. Findings. Network Design Lab; 2020; doi:10.32866/001c.12903
37.
Feyman Y, Bor J, Raifman J, Griffith KN. Effectiveness of COVID-19 shelter-in-place orders varied by state. PLOS ONE. 2020;15: e0245008. doi:10.1371/journal.pone.0245008
38.
Praharaj S, King D, Pettit C, Wentz E. Using aggregated mobility data to measure the effect of COVID-19 policies on mobility changes in sydney, london, phoenix, and pune. Findings. Network Design Lab; 2020; doi:10.32866/001c.17590
39.
Paez A. Using google community mobility reports to investigate the incidence of COVID-19 in the united states. Findings. 2020; doi:`https://doi.org/10.32866/001c.12976`
40.
Noland RB. Mobility and the effective reproduction rate of COVID-19. Journal of Transport & Health. 2021;20: 101016. doi:`https://doi.org/10.1016/j.jth.2021.101016`
41.
Bhadra A, Mukherjee A, Sarkar K. Impact of population density on covid-19 infected and mortality rate in india. Modeling Earth Systems and Environment. 2021;7: 623–629. doi:10.1007/s40808-020-00984-7
42.
Kadi N, Khelfaoui M. Population density, a factor in the spread of COVID-19 in algeria: Statistic study. Bulletin of the National Research Centre. 2020;44. doi:10.1186/s42269-020-00393-x
43.
Pequeno P, Mendel B, Rosa C, Bosholn M, Souza JL, Baccaro F, et al. Air transportation, population density and temperature predict the spread of COVID-19 in brazil. PeerJ. 2020;8: e9322. doi:10.7717/peerj.9322
44.
Fielding-Miller RK, Sundaram ME, Brouwer K. Social determinants of COVID-19 mortality at the county level. PLOS ONE. 2020;15: e0240151. doi:10.1371/journal.pone.0240151
45.
Roy S, Ghosh P. Factors affecting COVID-19 infected and death rates inform lockdown-related policymaking. PLOS ONE. 2020;15: e0241165. doi:10.1371/journal.pone.0241165
46.
Wong DWS, Li Y. Spreading of COVID-19: Density matters. PLOS ONE. 2020;15: e0242398. doi:10.1371/journal.pone.0242398
47.

Sun Z, Zhang H, Yang Y, Wan H, Wang Y. Impacts of geographic factors and population density on the COVID-19 spreading under the lockdown policies of china. Science of The Total Environment. 2020;746: 141347. doi:10.1016/j.scitotenv.2020.141347
48.
Skórka P, Grzywacz B, Moroń D, Lenda M. The macroecology of the COVID-19 pandemic in the anthropocene. PLOS ONE. 2020;15: e0236856. doi:10.1371/journal.pone.0236856
49.
White ER, Hébert-Dufresne L. State-level variation of initial COVID-19 dynamics in the united states. PLOS ONE. 2020;15: e0240648. doi:10.1371/journal.pone.0240648
50.
Khavarian-Garmsir AR, Sharifi A, Moradpour N. Are high-density districts more vulnerable to the COVID-19 pandemic? Sustainable Cities and Society. 2021;70: 102911. doi:10.1016/j.scs.2021.102911
51.
Hamidi S, Ewing R, Sabouri S. Longitudinal analyses of the relationship between development density and the COVID-19 morbidity and mortality rates: Early evidence from 1,165 metropolitan counties in the united states. Health & Place. 2020;64: 102378. doi:10.1016/j.healthplace.2020.102378
52.
Lee M, Zhao J, Sun Q, Pan Y, Zhou W, Xiong C, et al. Human mobility trends during the early stage of the COVID-19 pandemic in the united states. PLOS ONE. 2020;15: e0241468. doi:10.1371/journal.pone.0241468
53.
Amadu I, Ahinkorah BO, Afitiri A-R, Seidu A-A, Ameyaw EK, Hagan JE, et al. Assessing sub-regional-specific strengths of healthcare systems associated with COVID-19 prevalence, deaths and recoveries in africa. PLOS ONE. 2021;16: e0247274. doi:10.1371/journal.pone.0247274
54.
Feng Y, Li Q, Tong X, Wang R, Zhai S, Gao C, et al. Spatiotemporal spread pattern of the COVID-19 cases in china. PLOS ONE. 2020;15: e0244351. doi:10.1371/journal.pone.0244351
55.
Inbaraj LR, George CE, Chandrasingh S. Seroprevalence of COVID-19 infection in a rural district of south india: A population-based seroepidemiological study. PLOS ONE. 2021;16: e0249247. doi:10.1371/journal.pone.0249247
56.
Souris M, Gonzalez J-P. COVID-19: Spatial analysis of hospital case-fatality rate in france. PLOS ONE. 2020;15: e0243606. doi:10.1371/journal.pone.0243606
57.
Ahmad K, Erqou S, Shah N, Nazir U, Morrison AR, Choudhary G, et al. Association of poor housing conditions with COVID-19 incidence and mortality across US counties. PLOS ONE. 2020;15: e0241327. doi:10.1371/journal.pone.0241327
58.
Noury A, François A, Gergaud O, Garel A. How does COVID-19 affect electoral participation? Evidence from the french municipal elections. PLOS ONE. 2021;16: e0247026. doi:10.1371/journal.pone.0247026
59.
Wang F, Tan Z, Yu Z, Yao S, Guo C. Transmission and control pressure analysis of the COVID-19 epidemic situation using multisource spatio-temporal big data. PLOS ONE. 2021;16: e0249145. doi:10.1371/journal.pone.0249145
60.

Brandtner C, Bettencourt LMA, Berman MG, Stier AJ. Creatures of the state? Metropolitan counties compensated for state inaction in initial u.s. Response to COVID-19 pandemic. PLOS ONE. 2021;16: e0246249. doi:10.1371/journal.pone.0246249

61.

Stephens KE, Chernyavskiy P, Bruns DR. Impact of altitude on COVID-19 infection and death in the united states: A modeling and observational study. PLOS ONE. 2021;16: e0245055. doi:10.1371/journal.pone.0245055