

README for

“Spillovers and Imputation: Difference-in-Differences with Spatial Exposure”

Overview

This package contains all code and data required to replicate the results of the paper *“Spillovers and Imputation: Difference-in-Differences with Spatial Exposure.”* The paper develops and evaluates a spillover imputation method for difference-in-differences (DID) designs in which spatial spillovers contaminate standard control groups. The package includes two components: (i) a Monte Carlo simulation with a known data-generating process that benchmarks the proposed method against standard DID and ring-augmented DID, and (ii) an applied case that re-analyzes a road-paving randomized experiment in Colombia.

To replicate all results, a user must (1) obtain the restricted data files described in the Data Availability section, (2) place them in the `Data/` folder, (3) update the working directory path in `master.R`, and (4) run `master.R`. The script will install all required R packages and sequentially execute all analysis files, saving all tables and figures to the `Results/` folder.

Data Availability

Some data cannot be made publicly available.

Data Sources

The applied case in this paper uses data from the following unpublished working paper:

Dataset 1: Household survey data

- **Filename:** `ALSS_data.csv`
 - **Source:** González-Quintana et al. (IZA Discussion Paper No. 18082)
 - **URL:** <https://www.iza.org/publications/dp/18082/local-public-goods-and-property-tax-compliance-experimental-evidence-from-street-pavement>
 - **Access year:** 2025
 - **Variables used:** Household identifier (`id_hogar_06_09`), project cluster (`proyecto`), baseline and follow-up adjusted log home values (`adj_log_value_home_occ_b`, `adj_log_value_home_occ`), treatment assignment (`paved2`, `intent_to_treat2`), geographic coordinates (`lat`, `lon`), survey weights (`POND0609`)
 - **Access restrictions:** This dataset originates from an unpublished IZA Discussion Paper. The data are not yet publicly available through the IZA repository and access remains limited pending final publication. The authors of the original study have explicitly granted permission to use their data for this project.
-

Dataset 2: Street-level shapefile

- **Filename:** `Experimental_streets.shp` (and associated files: `.dbf`, `.prj`, `.shx`, `.sbn`, `.sbx`, `.cpg`)

- **Source:** González-Quintana et al. (IZA Discussion Paper No. 18082)
- **URL:** <https://www.iza.org/publications/dp/18082/local-public-goods-and-property-tax-compliance-experimental-evidence-from-street-pavement>
- **Access year:** 2025
- **Description:** Polyline shapefile of experimental streets in the study area, with attributes for paving year (`pav_year`) and paving status. Coordinate reference system: WGS84 (EPSG:4326).
- **Access restrictions:** Same restrictions as Dataset 1 above. Data access is pending publication of the source paper. Permission to use has been granted by the original authors.

Both datasets come from the same source. The corresponding author is Climent Quintana; email: C.Quintana-Domeque@exeter.ac.uk.

Dataset 3: Simulation data (generated by code)

No external data file is required for the simulation component. All data are generated internally by `Code/functions/simulate_spatial_panel.R` at run time using a fixed random seed.

Statement about Rights

- I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- The authors of the source paper (González-Quintana et al.) have provided explicit permission to use their data for this research project. Their data are not redistributed publicly within this package and are included for reproducibility verification purposes only.

Instructions for Replicators

1. **Obtain the restricted data.** Contact the authors of González-Quintana et al. (IZA DP No. 18082) to request access to `ALSS_data.csv` and the `Experimental_streets` shapefile. Also, their data will be publicly available from their website soon. Once access is granted, place the files in the following locations within the package:
 - `Data/ALSS_data.csv`
 - `Data/ALSS_data.xls`
 - `Data/Shapefiles/Experimental_streets.shp` (and all associated shapefile components)
2. **Update the working directory.** Open `master.R` and update **line 41** with the path to the root folder of this package on your machine:


```
setwd("C:/your/path/to/Replication_DID_Spillovers")
```

 Alternatively, open `master.R` in RStudio and use *Session > Set Working Directory > To Source File Location*, then comment out line 41.
3. **Install R.** The package was developed using R. Ensure R (version 4.0 or later) is installed. RStudio is recommended but not required.

4. **Run the master script.** Execute `master.R` in full (e.g., `source("master.R")` from the R console, or Ctrl+Shift+Enter in RStudio). The script will automatically install all required packages and run all analysis files in sequence.
5. **Check outputs.** All tables and figures will be saved to the `Results/` folder:
 - `Results/Tables/` — LaTeX table files (`.tex`)
 - `Results/Figures/` — Figure files (`.png`)

List of Exhibits

The provided code reproduces:

- ☒ All tables and figures in the paper
- ☒ All numbers provided in text in the paper

Exhibit	Output filename	Script	Notes
Figure 1	Figure1.tex	TikZ — embedded directly in LaTeX manuscript	Conceptual diagram of a six-unit spatial network illustrating exposure functions and spillover notation; found in Results/Figures/
Figure 2	Figure2.tex	TikZ — embedded directly in LaTeX manuscript	Diagram comparing distance-ring vs. neighbourhood-exchangeability exposure mappings; found in Results/Figures/
Figure 3	Figure3.tex	TikZ — embedded directly in LaTeX manuscript	Two-panel plot illustrating the DID setup with spillovers (theoretical groups vs. observed treated/control groups); found in Results/Figures/
Figure 4	Figure4.tex	TikZ — embedded directly in LaTeX manuscript	Four-panel decomposition theorem visualization showing all 2×2 DID combinations across treated/control exposure levels; found in Results/Figures/
Table 1	Table1.tex	Code/01_simulation.R	Found in Results/Tables/
Table 2	Table2.tex	Code/01_simulation.R	Found in Results/Tables/
Table 3	Table3.tex	Code/01_simulation.R	Found in Results/Tables/
Table 4	Table4.tex	Code/02_applied_case.R	Found in Results/Tables/; requires restricted data

Exhibit	Output filename	Script	Notes
Figure 5A	Fig_5-A.png	Code/01_simulation.R	Found in Results/Figures/; PNG, 32-bit color
Figure 5B	Fig_5-B.png	Code/01_simulation.R	Found in Results/Figures/; PNG, 32-bit color
Figure 6	Fig_6.png	Code/01_simulation.R	Found in Results/Figures/; PNG, 32-bit color
Figure 7	Fig_7.png	Code/01_simulation.R	Found in Results/Figures/; PNG, 32-bit color
Figure 8	Fig_8.png	Code/02_applied_case.R	Found in Results/Figures/; requires restricted data
Figure A10	FigA10_dynamic_effects.png	Code/01_simulation.R	Found in Results/Figures/; PNG, 32-bit color
Figure A11-A	Fig_A11-A.png	Code/03_appendix_figures.R	Found in Results/Figures/; requires restricted data and internet connection for OSM tiles
Figure A11-B	Fig_A11-b.png	Code/03_appendix_figures.R	Found in Results/Figures/; requires restricted data and internet connection for OSM tiles
Figure A11-C	Fig_A11-c.png	Code/03_appendix_figures.R	Found in Results/Figures/; requires restricted data and internet connection for OSM tiles
Figure A11-D	Fig_A11-d.png	Code/03_appendix_figures.R	Found in Results/Figures/; requires restricted data and internet connection for OSM tiles

Note on data requirements by exhibit:

- **Figures 1–4** are pure TikZ/LaTeX diagrams with no data dependency. They are self-contained .tex files included directly in the manuscript source and do not require running any R code.
- **Tables 1–3 and Figures 5–7 and A10** can be reproduced without the restricted data, as they rely entirely on simulated data generated by the code at run time.
- **Table 4 and Figures 8 and A11** require the restricted ALSS dataset and shapefiles from González-Quintana et al.

Figures A11-A through A11-D rely on OpenStreetMap background tiles fetched at run time via the `ggspatial / rosm` packages. An internet connection is required the first time these figures are produced; tiles are cached in `rosm.cache/` for subsequent runs.

Requirements

Software

Software	Version
R	4.0 or later (developed and tested on R 4.x)
RStudio	Recommended (any recent version)

R Packages

The following packages are automatically installed by `master.R` (lines 46–48). All are available on CRAN.

Package	Purpose
<code>rlang</code>	Language utilities
<code>dplyr</code>	Data manipulation
<code>tidyr</code>	Data reshaping
<code>fixest</code>	High-dimensional fixed-effects regressions
<code>kableExtra</code>	LaTeX table formatting
<code>ggplot2</code>	Figures
<code>sf</code>	Spatial data operations
<code>ggspatial</code>	Spatial layers and map tiles in <code>ggplot2</code>
<code>ggnewscale</code>	Multiple color/fill scales in <code>ggplot2</code>
<code>ggpmisc</code>	Regression equation annotations in <code>ggplot2</code>

Memory, Runtime, and Storage

- **Approximate run time:** Under 10 minutes on a standard laptop (most time is spent on the Monte Carlo simulation in `01_simulation.R`).
- **Storage:** The package as provided occupies approximately 50 MB, dominated by the figure outputs and the OSM tile cache. During a full run, no large intermediate files are written to disk.
- **Memory:** No special memory requirements; 8 GB RAM is sufficient.

Code Description

- **Results/Figures/Figure1.tex – Figure4.tex** — Conceptual TikZ/LaTeX figures (Figures 1–4 in the paper). These are standalone .tex files written by hand that are input-ed directly into the manuscript. They are not generated by any R script and require no data; they are ready to use as-is.
- **master.R** — Main entry point. Sets the working directory, installs and loads all required packages, sources the three helper functions, and sequentially calls the three analysis scripts.
- **Code/01_simulation.R** — Monte Carlo simulation of spatial spillovers under a DID design. Generates Tables 1–3 and Figures 5A, 5B, 6, 7, and A10.
- **Code/02_applied_case.R** — Applied analysis using the González-Quintana road-paving experiment. Computes distances to the nearest paved street, constructs a panel dataset, estimates

treatment and spillover effects using four methods (2SLS, naive DID, ring-augmented DID, and spillover imputation), and generates Table 4 and Figure 8.

- **Code/03_appendix_figures.R** — Produces spatial map figures for the appendix (Figure A11 panels A–D), showing aerial views of three example households with concentric buffer rings and paved/unpaved street overlays. Depends on objects created in 02_applied_case.R and must be run after it.
- **Code/functions/simulate_spatial_panel.R** — Data-generating function for the simulation. Creates a 100×100 spatial grid with treated, exposed, and never-exposed units across multiple time periods, and assigns outcomes based on unit and time fixed effects, a treatment effect, distance-decaying spillovers, and noise.
- **Code/functions/spill_imputation.R** — Core estimation function implementing the spillover imputation method. Uses never-exposed control units and pre-treatment observations to fit a unit- and time-fixed-effects model, then computes individual treatment and spillover effects as residuals.
- **Code/functions/plot_spill_estimates.R** — Utility function that plots dynamic treatment and spillover effect estimates with 95% confidence intervals across time periods.

Folder Structure

```

Replication_DID_Spillovers/
├── master.R
├── README.md
├── Code/
│   ├── 01_simulation.R
│   ├── 02_applied_case.R
│   ├── 03_appendix_figures.R
│   └── functions/
│       ├── simulate_spatial_panel.R
│       ├── spill_imputation.R
│       └── plot_spill_estimates.R
├── Data/
│   ├── ALSS_data.csv          ← restricted; see Data Availability
│   ├── ALSS_data.xls          ← restricted; see Data Availability
│   └── Shapefiles/
│       ├── Experimental_streets.shp
│       ├── Experimental_streets.dbf
│       ├── Experimental_streets.prj
│       ├── Experimental_streets.shx
│       ├── Experimental_streets.sbn
│       ├── Experimental_streets.sbx
│       └── Experimental_streets.cpg
├── Results/
│   └── Figures/
│       ├── Figure1.tex          ← TikZ figure, embedded directly in manuscript
│       ├── Figure2.tex          ← TikZ figure, embedded directly in manuscript
│       ├── Figure3.tex          ← TikZ figure, embedded directly in manuscript
│       ├── Figure4.tex          ← TikZ figure, embedded directly in manuscript
│       ├── Fig_5-A.png
│       ├── Fig_5-B.png
│       ├── Fig_6.png
│       ├── Fig_7.png
│       ├── Fig_8.png
│       ├── FigA10_dynamic_effects.png
│       ├── Fig_A11-A.png
│       ├── Fig_A11-b.png
│       └── Fig_A11-c.png

```

```
├── Fig_A11-d.png
├── Tables/
│   ├── Table1.tex
│   ├── Table2.tex
│   ├── Table3.tex
│   └── Table4.tex
└── rosm.cache/
```

← OSM tile cache (auto-generated)