

Thailand Cities of the Future – SQGE Reproducibility Package
=====

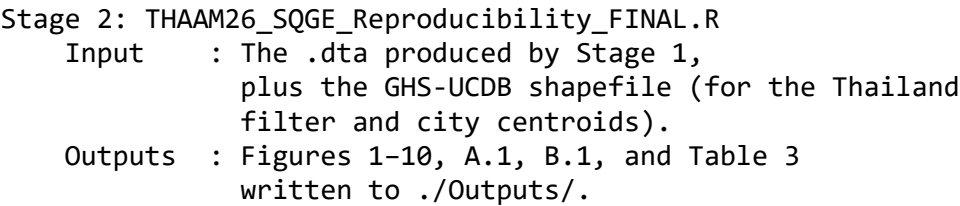
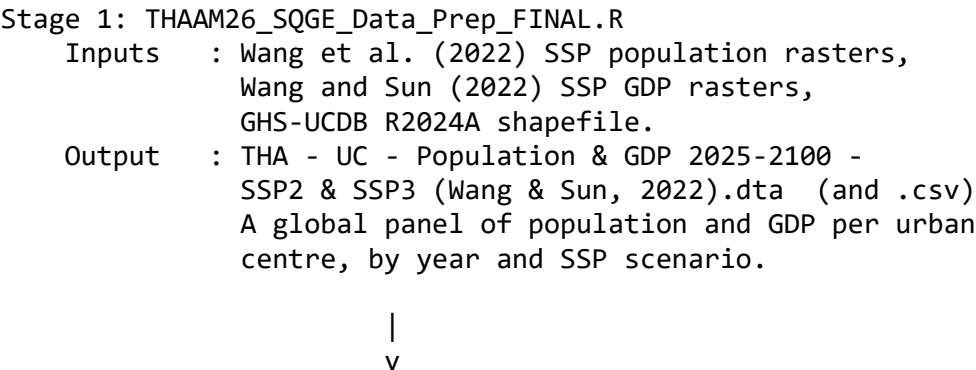
Working paper title. Thailand Cities of the Future: Spatial Quantitative General Equilibrium Analysis of Urban Development Scenarios, 2025–2050.

Authors. Putu Sanjiwacika Wibisana, Steven Louis Rubinyi.

Submission. World Bank Policy Research Working Paper (PWRP) series.

This package reproduces every figure and the headline table reported in the paper from raw projection rasters and GHS-UCDB urban-centre boundaries. The pipeline runs in two stages, executed in order: first a data-preparation script that builds the city-year panel from the upstream raster products, then a reproducibility script that calibrates the SQGE model and writes the figures and Table 3.

Workflow at a glance



Stage 1 is heavy and runs only when raw rasters change. Stage 2 is the reproducibility entry point and runs against the .dta from Stage 1.

Folder structure

PWRP Reproducibility Package/		
THAAM26_SQGE_Data_Prep_FINAL.R		Stage 1 script
THAAM26_SQGE_Reproducibility_FINAL.R		Stage 2 script

README.txt	this file
Data/	inputs and the
	Stage 1 output
GHS_UCDB_GLOBE_R2024A_UC_Boundaries.shp	
(plus the rest of the GHS-UCDB shapefile sidecar files)	
Projected SSP Population (Wang 2022)/	
SSP2_2025.tif, SSP2_2030.tif, ...	
SSP3_2025.tif, SSP3_2030.tif, ...	
FIXED/	produced by Stage 1
Global (Wang & Sun, 2022)/	
SSP Projection/	
GDP2025_ssp2.tif, GDP2030_ssp2.tif, ...	
GDP2025_ssp3.tif, GDP2030_ssp3.tif, ...	
THA - UC - Population & GDP 2025-2100 - SSP2 & SSP3	
(Wang & Sun, 2022).dta	Stage 1 output
THA - UC - Population & GDP 2025-2100 - SSP2 & SSP3	
(Wang & Sun, 2022).csv	Stage 1 output
Simulation.csv	created on-the-fly
	by Stage 2
Outputs/	Stage 2 figures and table
Intermediates/	cached .rds files

Quick start

1. Place the raw inputs in the Data/ subfolder as shown above:
 - a. Wang et al. (2022) SSP population rasters in "Data/Projected SSP Population (Wang 2022)/"
 - b. Wang and Sun (2022) SSP GDP rasters in "Data/Global (Wang & Sun, 2022)/SSP Projection/"
 - c. GHS-UCDB R2024A shapefile and sidecars in "Data/"
 (See "References" at the foot of this README for the source URLs.)

2. Open R or RStudio. In each of the two scripts, edit the setwd() line near the top so it points to your local copy of this package.

3. Run Stage 1 (data preparation):

```
source("THAAM26_SQGE_Data_Prep_FINAL.R")
```

This builds the global UC × year × scenario panel from the rasters and saves it as "THA - UC - Population & GDP 2025-2100 - SSP2 & SSP3 (Wang & Sun, 2022).dta" (plus a parallel .csv) in your data directory. Confirm the file lands in "Data/" before continuing.

4. Run Stage 2 (reproducibility):

```
source("THAAM26_SQGE_Reproducibility_FINAL.R")
```

The Stage 2 script reads the .dta from Stage 1, filters to Thailand and the SSP2 scenario, attaches city centroids from the shapefile, writes a working file "Data/Simulation.csv" internally, then calibrates the SQGE model and writes the figures and Table 3 to ./Outputs/.

Outputs produced

All thirteen figures and the headline table from the working paper are written by Stage 2 to ./Outputs/.

Figure 1.svg	National urban GDP per capita by scenario, 2025–2050 (2017 USD PPP)
Figure 2.svg	GDP per capita gain above no-investment baseline by scenario, 2030–2050 (pp)
Figure 3.svg	GDP by city tier, indexed to 2025 = 100
Figure 4.svg	GDP per capita by city tier, indexed to 2025 = 100
Figure 5.svg	Population by city tier, indexed to 2025 = 100
Figure 6.svg	GDPpc trajectories by individual city, indexed to 2025 = 100
Figure 7.svg	Gini coefficient of city-level GDP by scenario, 2025–2050
Figure 8.svg	Tipping-point curve: 2050 GDPpc gain by envelope size (Scenarios A and C vs the 50/50 baseline)
Figure 9.svg	Envelope × split heatmap of 2050 GDP per capita gains
Figure 10.svg	Year of crossover between 80/20 and 20/80 portfolios, by envelope size
Figure A.1.svg	Sensitivity of GDP per capita gain to agglomeration elasticity (γ)
Figure B.1.svg	Sensitivity of GDP per capita gain to worker reallocation elasticity (ϵ)
Table 3.xlsx	National GDP and GDP per capita gains by scenario, 2030–2050

Data Availability Statement

Accessibility category: All data used in this package are PUBLICLY AVAILABLE under open-data licenses.

The analysis draws on three external datasets, each obtained from a permanent, citable repository under an open-access licence. None of the raw input data is redistributed in this package; users must obtain the input files directly from the original sources listed below.

1) GHS-UCDB R2024A – urban-centre boundaries

Filename(s) used in this package:

GHS_UCDB_GLOBE_R2024A_UC_Boundaries.shp
GHS_UCDB_GLOBE_R2024A_UC_Boundaries.dbf
GHS_UCDB_GLOBE_R2024A_UC_Boundaries.prj
GHS_UCDB_GLOBE_R2024A_UC_Boundaries.shx
GHS_UCDB_GLOBE_R2024A_UC_Boundaries.cpg

Source:

European Commission, Joint Research Centre (JRC)

URL:

https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/GHSL/GHS_UCDB_GLOBE_R2024A/GHS_UCDB_GLOBE_R2024A/V1-0/

Access date:

January 2026

Full citation:

European Commission, Joint Research Centre (2024). GHS-UCDB
R2024A – GHS Urban Centre Database 2025. [Dataset.]
DOI: 10.2905/1a338be6-7eaf-480c-9664-3a8ade88cbcd.

Licence:

Open data. Reuse permitted under the JRC reuse policy with proper acknowledgment of the source ("© European Union, 1995–2025. Reuse of this data is authorised with proper acknowledgment of the source.").

Access instructions:

Download the "GHS-UCDB R2024A – GHS Urban Centre Database 2025" dataset from the URL above. Within the global shapefile, all 11,422 urban centres are retained; the Stage 2 reproducibility script filters to Thailand using the GC_CNT_G == "Thailand" attribute. Place the shapefile and all its sidecar files (.dbf / .prj / .shx / .cpg) in the Data/ folder.

2) Wang and Sun (2022) – Global gridded GDP under historical and future scenarios

Filename(s) used in this package:

SSP2.7z (contains GDP{YEAR}_ssp2.tif for 2025–2100)
SSP3.7z (contains GDP{YEAR}_ssp3.tif for 2025–2100)

After extraction, the per-year per-scenario rasters used are:
GDP2025_ssp2.tif, GDP2030_ssp2.tif, ..., GDP2100_ssp2.tif

Source:

Zenodo (record managed by the authors).

URL:

<https://zenodo.org/records/7898409>

Access date:

January 2026

Full citation:

Wang, T. and Sun, F. (2022). Global gridded GDP under the historical and future scenarios. [Dataset.] Zenodo.
DOI: 10.5281/zenodo.7898409.

Licence:

Creative Commons Attribution 4.0 International (CC BY 4.0).
Free reuse with attribution.

Access instructions:

Download SSP2.7z and SSP3.7z from the Zenodo record above (approximately 7.2 GB each). Extract the per-year GDP rasters and place them in:

Data/Global (Wang & Sun, 2022)/SSP Projection/
The 16 SSP2 and 16 SSP3 yearly rasters (2025–2100 at five-year intervals) are the files actually consumed by Stage 1. All other archives in the Zenodo record (SSP1, SSP4, SSP5, the historical 2000–2020 files, and the 0.25-degree variant) are not used. Values are stored in 2005 USD PPP; Stage 2 rescales them to 2017 USD PPP using the World Bank ICP 2017 conversion factor (0.91525) before any analysis.

3) Wang, Meng and Long (2022) – Projected 1 km-grid SSP population

Filename(s) used in this package:

SSP2_2025.tif, SSP2_2030.tif, ..., SSP2_2100.tif

Source:

Figshare (record managed by the authors).

URL:

https://figshare.com/articles/dataset/Projecting_1_km-grid_population_distributions_from_2020_to_2100_globally_under_shared_socioeconomic_pathways/19608594

Access date:

January 2026

Full citation:

Wang, X., Meng, X. and Long, Y. (2022). Projecting 1 km-grid population distributions from 2020 to 2100 globally under shared socioeconomic pathways. *Scientific Data*, 9, 563.
Dataset DOI: 10.6084/m9.figshare.19608594.
Paper DOI: 10.1038/s41597-022-01675-x.

Licence:

Creative Commons Attribution 4.0 International (CC BY 4.0).
Free reuse with attribution.

Access instructions:

Download the SSP2 and SSP3 yearly population rasters from the Figshare record above. The 16 SSP2 and 16 SSP3 yearly rasters (2025–2100 at five-year intervals) are the files actually consumed by Stage 1. Place them in:
Data/Projected SSP Population (Wang 2022)/
Some rasters in this archive carry implausible extent metadata and an inherited N-S offset; Stage 1 detects this automatically and writes corrected rasters into a FIXED/ subfolder before extracting per-UC totals. Already-corrected files are skipped on subsequent runs.

Rights and permissions

The authors had open, lawful access to all three datasets above through the URLs listed, on the access dates indicated, under the respective open-data licences (CC BY 4.0 for the Wang and Sun GDP dataset and the Wang, Meng and Long population dataset, and the JRC reuse policy for the GHS-UCDB R2024A boundaries). All three licences permit reuse with attribution.

This package does NOT redistribute any of the raw input data. Users must obtain the input files directly from the original repositories listed above. The full data citations and access URLs are provided so that the input data can be reconstructed by any third party. All three repositories are permanent, citable archives with persistent identifiers (DOIs).

No personally identifiable information is contained in any of the input data; all three datasets are aggregate gridded products at 1 km² resolution and aggregate polygon-level attributes for global urban centres.

Stage 1 inputs and output (data preparation)

Stage 1 takes three families of input from upstream data products:

1. Wang et al. (2022) SSP population rasters.
Files of the form "SSP2_2025.tif", "SSP3_2030.tif", and so on, covering 2025 to 2100 at five-year intervals for SSP2 and SSP3. Stage 1 includes a sub-step that rebuilds the geographic extent of these rasters and applies a -6 degree latitude shift to align them with the rest of the inputs; the corrected rasters are written to a "FIXED" subfolder. Rasters already fixed are skipped on re-runs.
2. Wang and Sun (2022) SSP GDP rasters.
Files of the form "GDP2025_ssp2.tif", "GDP2030_ssp3.tif", and so on, distributed in 2005 USD PPP. No extent correction is needed.
3. GHS-UCDB R2024A shapefile.
The "GHS_UCDB_GLOBE_R2024A_UC_Boundaries.shp" file together with its sidecar .dbf / .prj / .shx / .cpg files.

Stage 1 produces a single panel file:

THA - UC - Population & GDP 2025-2100 - SSP2 & SSP3
(Wang & Sun, 2022).dta
(and the matching .csv)

Columns:

ID_UC_G	integer	GHS-UCDB urban-centre identifier
year	integer	2025, 2030, ..., 2100
scenario	string	"SSP2" or "SSP3"
gdp_sum	numeric	Total UC GDP, 2005 USD PPP
pop_sum	numeric	Total UC population
lon	numeric	Centroid longitude (WGS84)
lat	numeric	Centroid latitude (WGS84)
area_km2	numeric	UC area in square kilometres
popdens	numeric	pop_sum / area_km2
urban	logical	popdens > 1500 (DEGURBA core flag)

Stage 1 writes by default to the directory pointed to by the "gdData" global variable. Edit that line near the top of the script to redirect output if needed. The Stage 2 script expects the .dta in the working directory's "Data/" subfolder, so confirm the file is there before running Stage 2.

Stage 2 input

Stage 2 reads the .dta produced by Stage 1, plus the GHS-UCDB shapefile used to filter the global panel to Thailand and to attach centroids. The internal working file written by Stage 2 is

Data/Simulation.csv

with the schema:

ID_UC_G	integer	GHS-UCDB urban-centre identifier
city_name	string	Inherited from the shapefile attributes, mapped from name fields where available
year	integer	2025, 2030, ..., 2050 (Stage 2 truncates to the model horizon)
gdp_sum	numeric	Rescaled internally to 2017 USD PPP using 0.91525 (World Bank ICP 2017)
pop_sum	numeric	
area_km2	numeric	
lon	numeric	
lat	numeric	

GDP values are converted from 2005 USD PPP to 2017 USD PPP inside Stage 2 using the conversion factor 0.91525.

Bangkok and secondary cities. Bangkok is identified at calibration time as the Thai urban centre with the largest 2025 GDP. The eleven designated secondary cities are matched case-insensitively against the city_name field. If any of the following names is missing from the panel, Stage 2 will warn:

Chon Buri, Khon Kaen, Nakhon Pathom, Pattaya, Songkhla,
Chiang Mai, Hat Yai, Lop Buri, Nakhon Ratchasima, Phuket,
Ubon Ratchathani

Runtime expectations

Stage 1 (data preparation):

Raster extent fix (one-off per raster):
a few minutes per raster on first run; later runs skip already-fixed rasters and complete in seconds.

Population extraction (2 scenarios x 16 years x ~11 422 UCs):
20 to 40 minutes total.

GDP extraction (2 scenarios x 16 years x ~11 422 UCs):
20 to 40 minutes total.

Total Stage 1 cold run: roughly 45 to 90 minutes.

Stage 2 (reproducibility):

Cold run (no caches): 30 to 60 minutes on a typical laptop.
Stage-2 dynamic calibration and the epsilon sweep dominate.
Subsequent runs reuse the .rds caches in "Intermediates/"
and complete in 5 to 15 minutes.

To force a full recalibration, set `FORCE_RECALIBRATE <- TRUE` in
the CONFIG block at the top of the Stage 2 script.

Required R packages

Stage 1 requires:

`sf`, `terra`, `raster`, `exactextractr`, `dplyr`, `readr`, `haven`, `purrr`, `tidyr`

Stage 2 requires:

`sf`, `ggplot2`, `tidyverse` (incl. `dplyr`, `tidyr`, `readr`, `purrr`, `stringr`),
`haven`, `nloptr`, `scales`, `ggrepel`, `viridis`, `patchwork`, `openxlsx`,
`svglite`

Stage 2 will auto-install any missing CRAN packages on first run.
Stage 1 expects the spatial-package stack (`sf`, `terra`, `raster`,
`exactextractr`) to be installed in advance, since these depend on
external system libraries (GDAL, GEOS, PROJ) that R cannot install
on its own.

Tested under R 4.3.2 and R 4.4.0 on Windows 11 and Ubuntu 22.04 with
package versions broadly in the range tidyverse 2.0.0, nloptr 2.0.3,
scales 1.3.0, ggrepel 0.9.5, viridis 0.6.5, patchwork 1.2.0,
openxlsx 4.2.5, svglite 2.1.3, sf 1.0.16, terra 1.7.78,
exactextractr 0.10.0.

Model parameters

The central-case calibration uses the following structural parameters.
All are defined in the PAR list near the top of the Stage 2 script.

Parameter	Symbol	Value	Description
Migration elasticity (Fréchet shape) responsiveness	ϵ	1.4	Worker reallocation
Housing expenditure share share in utility	α	0.25	Cobb-Douglas housing
Capital income share	β	0.40	Cobb-Douglas capital

share in production				
Agglomeration elasticity	γ	0.11	TFP response to	
density (Melo, Graham				
			and Noland 2009	
central range)				
Congestion elasticity	μ	0.30	Cost-of-living	
response to density				
Trade cost decay	ν	0.001	Iceberg trade cost per	
km				
Market access elasticity	ξ	0.03	TFP response to market	
access				
Housing supply elasticity	η	1.50	Housing supply	
response to rent				
Capital depreciation rate (annual)	δ_K	0.05	Perpetual inventory	
depreciation				
Investment share of GDP (annual)	s_{inv}	0.25	Thailand average	
Amenity growth rate coefficient				
(log-saturating)	ϕ_{log}	0.012	Per-year	
Amenity per-period cap	B_{max}	1.15	Max 15 percent growth	
per 5-year step				
Time step	dt	5	Years per simulation	
period				

The central-case scenario design is:

Scenario	Bangkok share	$\delta\Omega$ Bangkok	$\delta\Omega$ per secondary city
No Investment	n/a	0.000	0.000
Scenario A	80%	+2.00%	+2.70% (×11 cities)
Scenario B	50%	+1.25%	+6.75% (×11 cities)
Scenario C	20%	+0.50%	+10.80% (×11 cities)

The envelope ratio (secondary shock per city / Bangkok shock) is fixed at 5.4 = 0.135 / 0.025. This ratio ensures that the aggregate productivity envelope is constant across scenarios at a given envelope size.

The envelope-by-split grid sweeps Bangkok-equivalent shock from 1 to 4 percent (in 0.25-percentage-point steps) and Bangkok share from 0.20 to 0.80 (in 0.10 steps), producing 13 × 7 = 91 grid cells.

The gamma sensitivity sweeps γ across {0.025, 0.04, 0.08, 0.11, 0.15, 0.20, 0.30}.

The epsilon sensitivity sweeps ϵ across {0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0}. The lower five values lie within the model's reliable solving range. At $\epsilon \geq 1.8$ the inner equilibrium solver returns a pathological corner equilibrium; these values are included for completeness and flagged in the paper.

Reproducibility notes

Hardware sensitivity of the calibration. The Stage-2 dynamic calibration minimises a sum-of-squared-errors objective with the gradient-free BOBYQA optimiser (nloptr, NLOPT_LN_BOBYQA). Gradient-free optimisers of this class follow a search path whose floating-point arithmetic is sensitive to CPU microarchitecture, vector-instruction width (for example AVX-512), and multi-socket parallelism. As a result, running the calibration from scratch on different hardware can terminate at a marginally different parameter vector and shift the headline gain figures by up to approximately 0.01 percentage points. To our knowledge these differences are numerical, not substantive, and do not affect any finding, ordering, or conclusion in the paper.

To eliminate this source of variation, the package ships the pre-computed calibration in Intermediates/calibration.rds. When this file is present, Stage 2 detects it, bypasses the optimiser, and reads the frozen calibration directly, so the calibration step is identical on any machine. A small residual difference in the final decimal place may still arise from the live equilibrium simulation that consumes the calibration, since floating-point summation order in the aggregation is itself mildly architecture-dependent. To reproduce the paper's exhibits exactly as published, run Stage 2 with the shipped Intermediates/ caches in place rather than forcing a full recalibration.

Setting `FORCE_RECALIBRATE <- TRUE` in the Stage 2 CONFIG block clears every cache and recomputes the model from scratch, which will reintroduce the hardware-dependent drift described above.

Re-running Stage 1 is only necessary when the upstream raster inputs or the GHS-UCDB shapefile change. If you have the Stage 1 output .dta file from a prior run, you can skip directly to Stage 2.

Caveats

The model is calibrated to SSP2 baseline projections. Sensitivity to alternative SSPs is not in this package.

All scenario shocks are applied as one-time multiplicative uplifts to the urban efficiency shifter Ω in 2030, and persist through 2050.

The model abstracts from intra-city transport networks; congestion enters as a city-level density penalty.

The eleven secondary cities are treated symmetrically within each scenario (same per-city shock). Heterogeneity across the secondary network is examined descriptively in Section 5.2 of the paper but is not endogenised.

Contact

For questions about the model, the data, or this reproducibility package, contact the corresponding author through the World Bank.

References

Allen, T. and Arkolakis, C. (2014). Trade and the Topography of the Spatial Economy. *Quarterly Journal of Economics*, 129(3), 1085–1140.

Desmet, K. and Rossi-Hansberg, E. (2014). Spatial Development. *American Economic Review*, 104(4), 1211–1243.

European Commission, Joint Research Centre (2024). Global Human Settlement Urban Centre Database (GHS-UCDB), Release R2024A.
Available at:
<https://data.jrc.ec.europa.eu/dataset/1a338be6-7eaf-480c-9664-3a8ade88cbcd>
Accessed: January 2026.

Melo, P. C., Graham, D. J. and Noland, R. B. (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39(3), 332–342.

Redding, S. J. and Rossi-Hansberg, E. (2017). Quantitative Spatial Economics. *Annual Review of Economics*, 9, 21–58.

Wang, T. and Sun, F. (2022). Global gridded GDP under the historical and future scenarios (SSP1–SSP5). Dataset, Zenodo.
Available at:
<https://zenodo.org/records/7898409>
Accessed: January 2026.

Wang, X., Meng, X. and Long, Y. (2022). Projecting 1 km-grid population distributions from 2020 to 2100 globally under shared socioeconomic pathways. *Scientific Data*, 9, 563. Dataset, Figshare.
Available at:

https://figshare.com/articles/dataset/Projecting_1_km-grid_population_distributions_from_2020_to_2100_globally_under_shared_socioeconomic_pathways/19608594
Accessed: January 2026.