Revisiting the value of public transport: An empirical study drawing on big data and open-source software

Anthropolis Seminar - March 2022

J Rafael Verduzco Torres
Urban Big Data Centre – Department of Urban Studies
University of Glasgow
February 2022

The role of public transport systems

- Public transport (PT) is key in the urban context for, at least:
  - Mitigating global and local pollutants
  - Addressing social justice concerns
  - Making efficient use of resources
- However, PT infrastructure is very costly
- Not only initial costs, but also operation & maintenance costs are burdensome (Murakami, 2012).

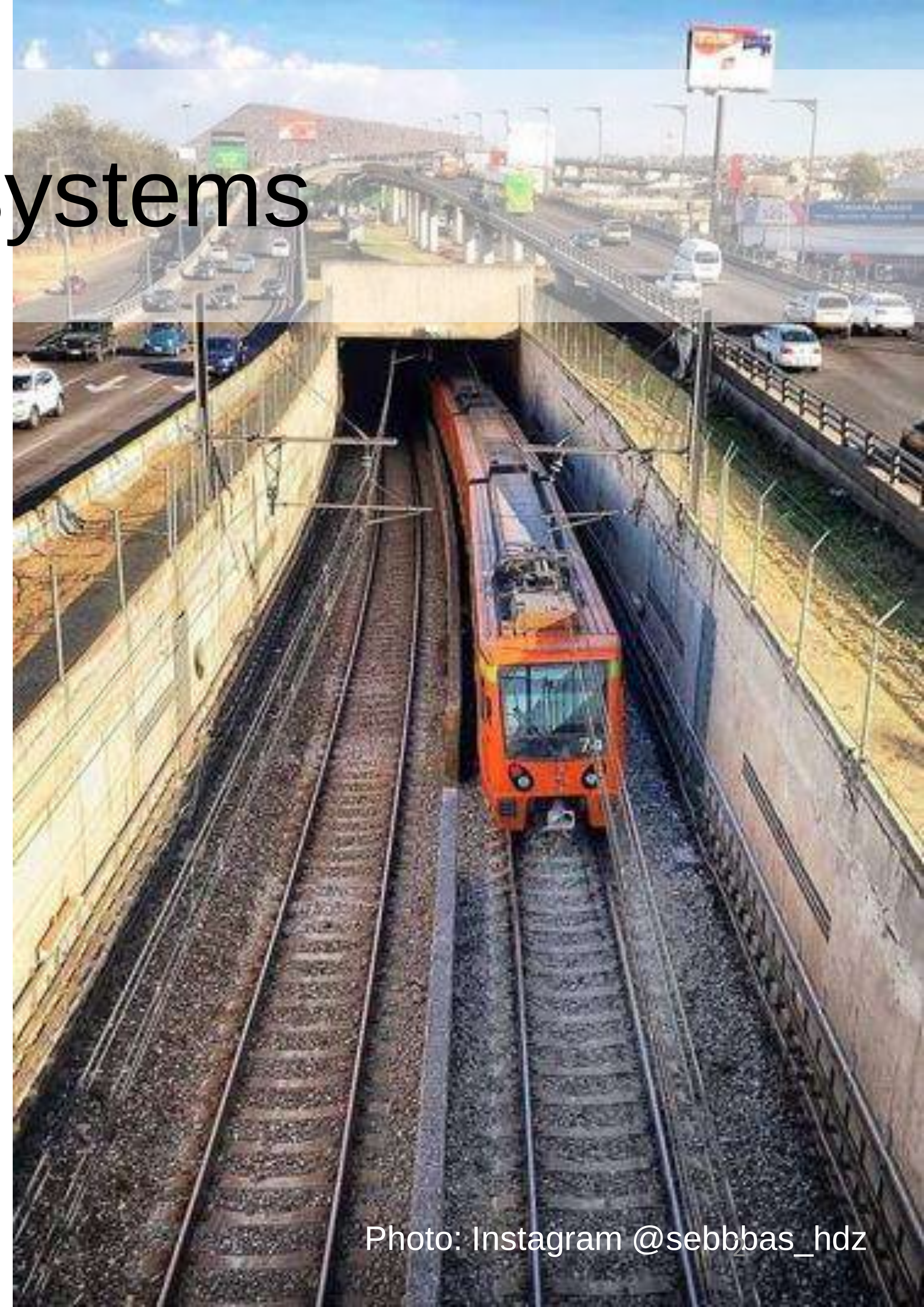


Photo: Instagram @sebbbas_hdz

Economic benefits

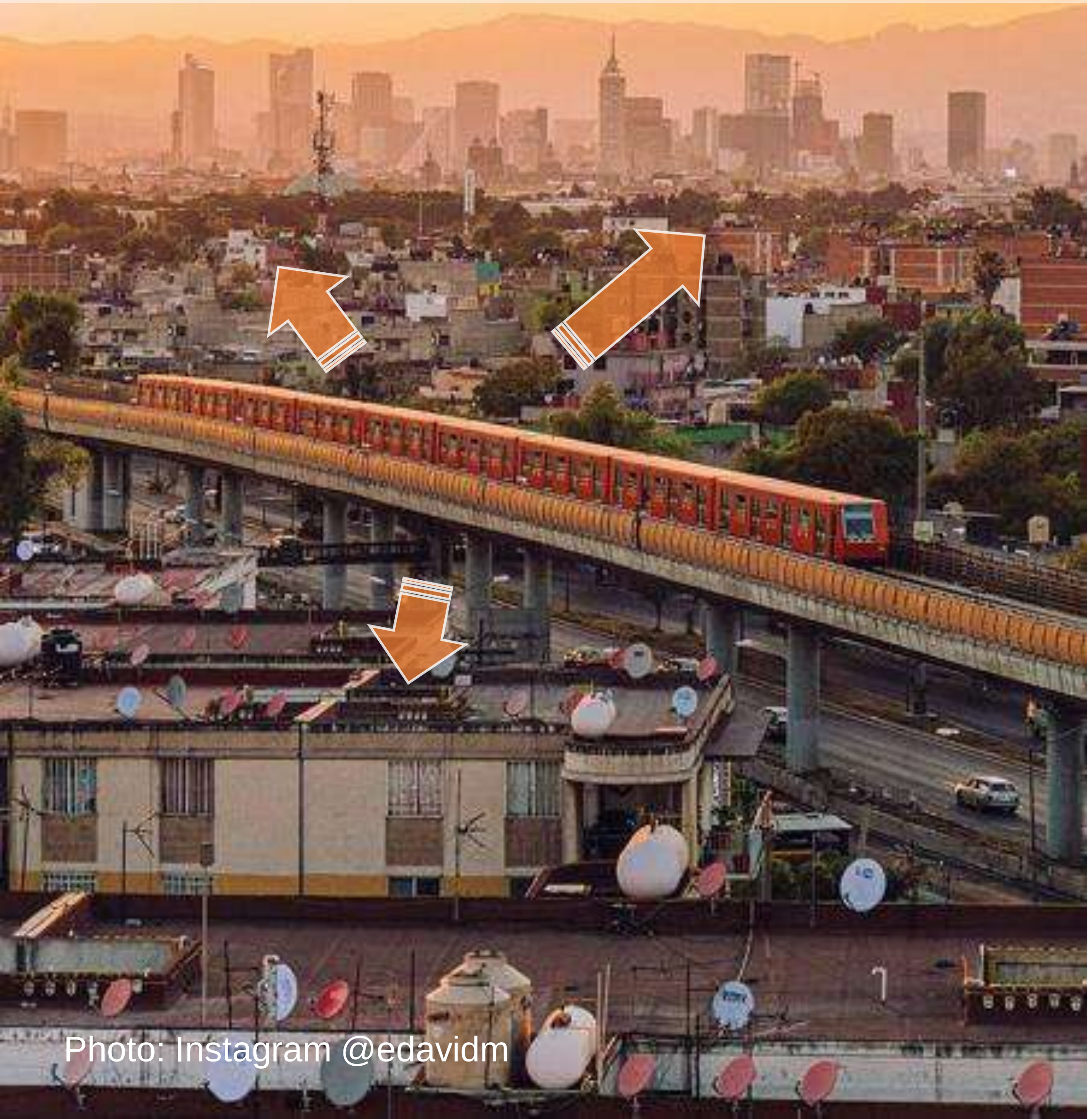


Photo: Instagram @edavidm

While governments are constrained, PT generates *special* and *direct* benefits that induce land value uplifts, e.g.:

- London, UK, 2.4 times (TfL, 2017);
- Perth, Australia, 0.6-1.3 times (McIntosh, 2015)
- Shanghai, China – 50% of economic benefits go to real estate value (Liu et al., 2018).

Mobilizing the economic benefits

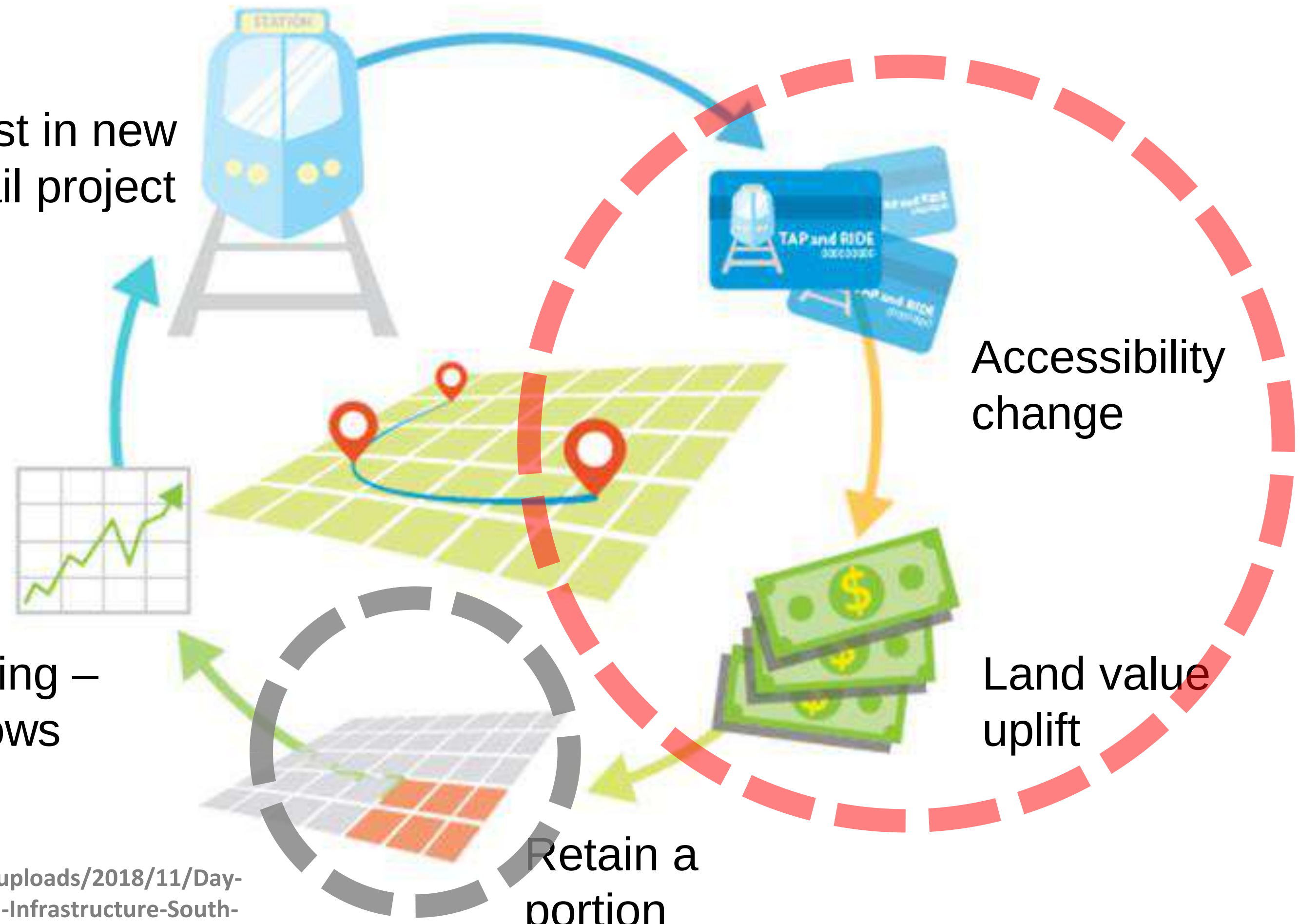
Sharing the economic benefits between landowners and the city can help to expand and maintain PT systems:

- Emblematic: Hong Kong (R+P) and Tokyo, Japan (land readjustment);[1]
- Crossrail (London), BRS 32% (£4.7 bn);[2]
- Ext. line 7 (New York, U.S.), 98% (US$2.4 bn);[2]
- Ouro Metro Line (Sao Paulo, Brazil), US$150 M.[3]

References: 1) Suzuki et al. (2015); 2)Salon et al. (2019); 3) Smolka and Maleronka (2018).

How?: Land value capture (LVC)

The conceptual cycle of value capture



Source: http://www.iifclprojects.com/wp-content/uploads/2018/11/Day-2_P3_Kathleen-Farrin_Implementing-Urban-Infrastructure-South-Asia-PPT.pdf

Research objective and questions

Objective:

Identify the extent of the potential of LVC for financing public transport drawing on big data and open-source resources in the case of Greater Mexico City.

The specific research questions (RQ) are:

- *How is accessibility to employment enabled by the main public transport network (MPTN)?*
- *What is the the willingness to pay for the accessibility generated by the main public transport network in the residential land market between 2009 and 2019?*

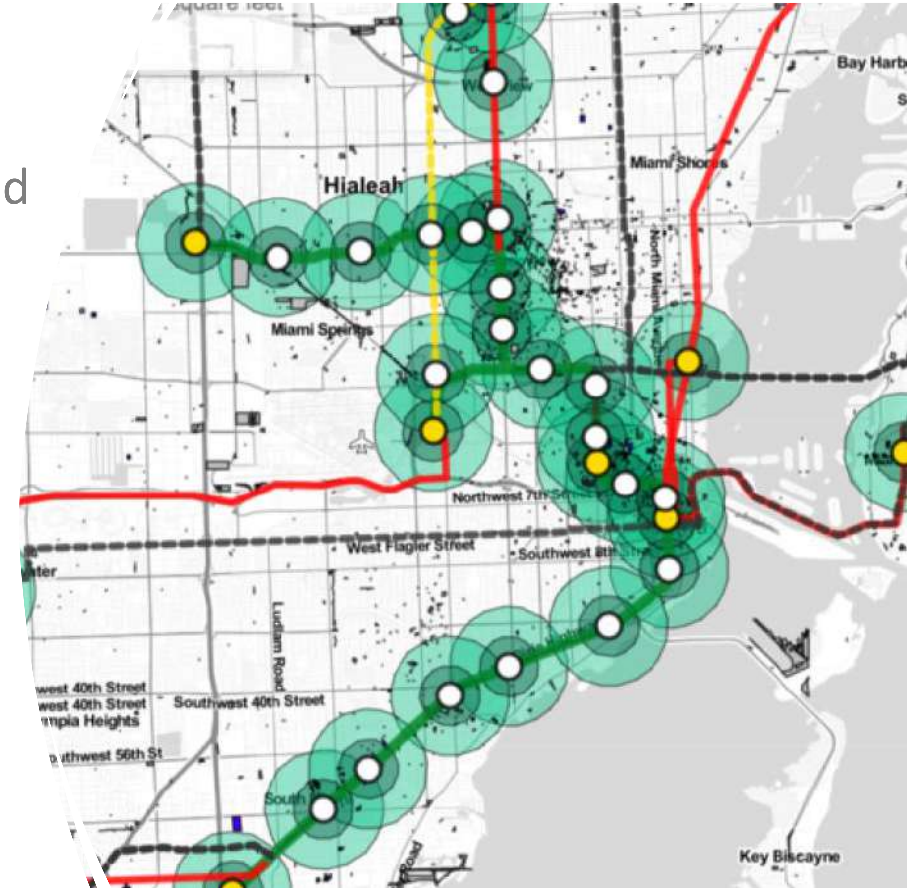
The assessment of value uplift

- *Hedonic price model:* Based on observed consumer's behaviour (implicit value).

$$P_{it} = f(z_{sit}, z_{lit})$$

- Neighbourhood
- Accessibility

- Very often the accessibility generated by PT is represented by oversimplified proxies (Higgins, 2016), e.g.:
    - Distance to closest station;
    - Number of stations within X radius;
    - Binary variables.
- These measures may consider where a journey start, but not the "effective" level of service.



Source: https://www.govtech.com/civic/mapping-tool-ids-potential-land-for-affordable-housing.html

Place-based accessibility measures

Opportunities at destinations

Origin

$$A_{ik} = \sum_{j=1}^{J} g(W_{jk})\, f(t_{ij})$$

Spatial links/ impedance

- Example: total number of jobs that can be reached within 60 minutes by public transport (*cumulative*)

Impedance functions in accessibility measures

$$A_{ik} = \sum_{j=1}^{J} g(W_{jk})\, f(t_{ij})$$

Cumulative rectangular

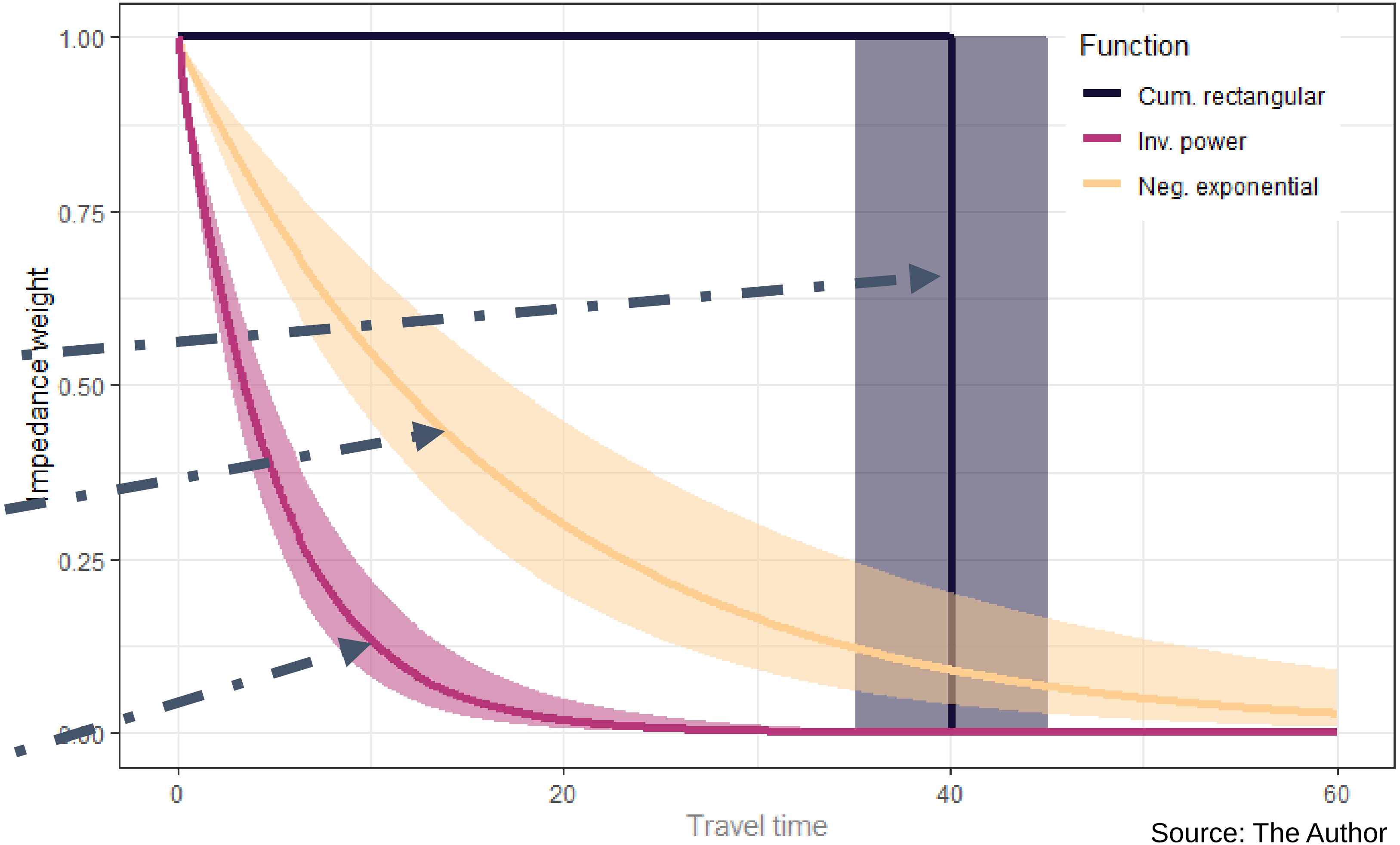
$$f(t_{ij}) = \begin{cases} 1 & \text{for } t_{ij} \leq \bar{t} \\ 0 & \text{otherwise} \end{cases}$$

Negative exponential

$$f(t_{ij}) = e^{(-\beta t_{ij})}$$

Inverse power

$$f(t_{ij}) = \begin{cases} 1 & \text{for } t_{ij} < 1 \\ t_{ij}^{-\beta} & \text{otherwise} \end{cases}$$



Source: The Author

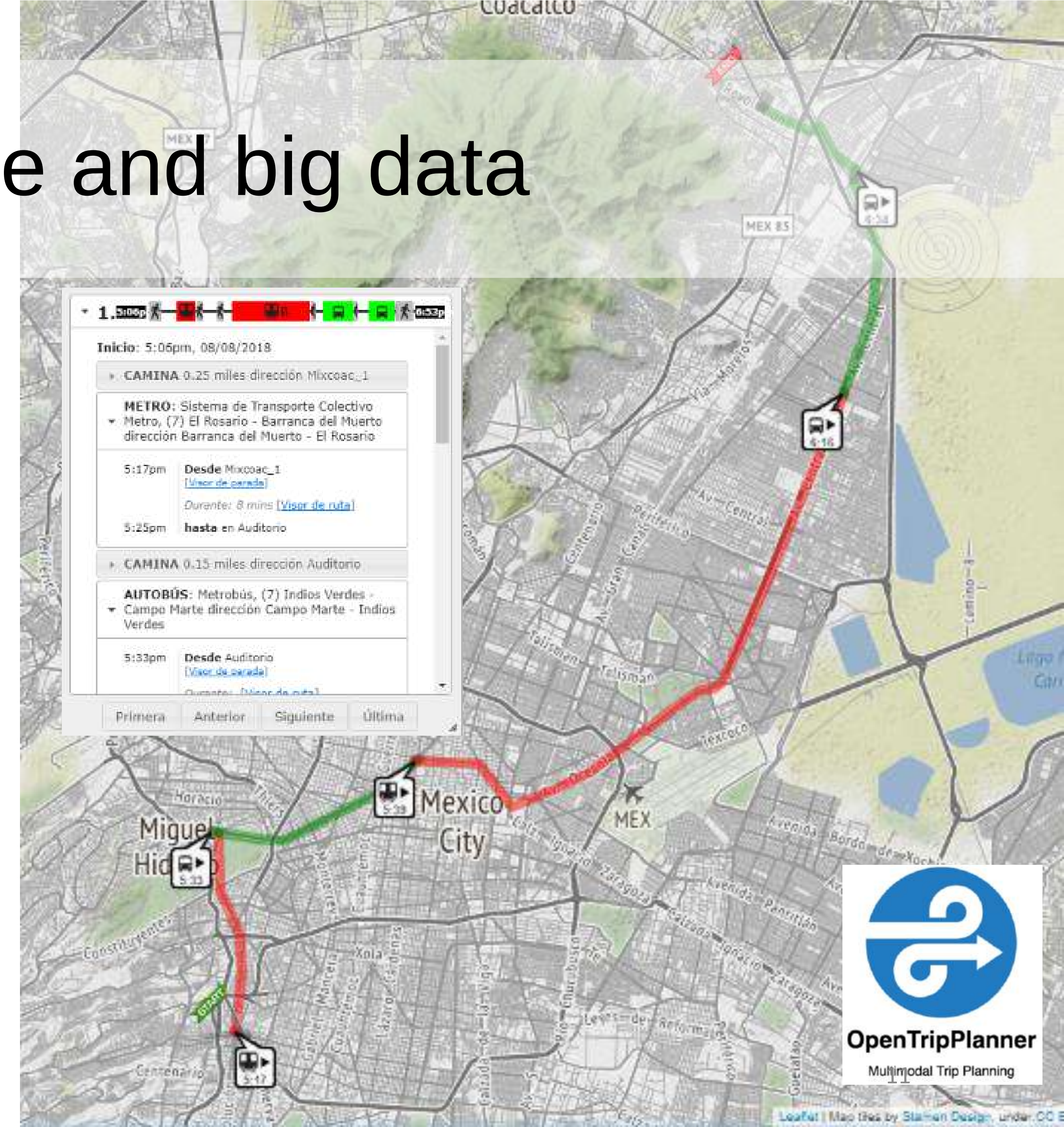
Place-based accessibility measures

- Advantages, considers:
  - The PT network
  - Land use (location of destinations), and
  - People’s preferences (potentially more)

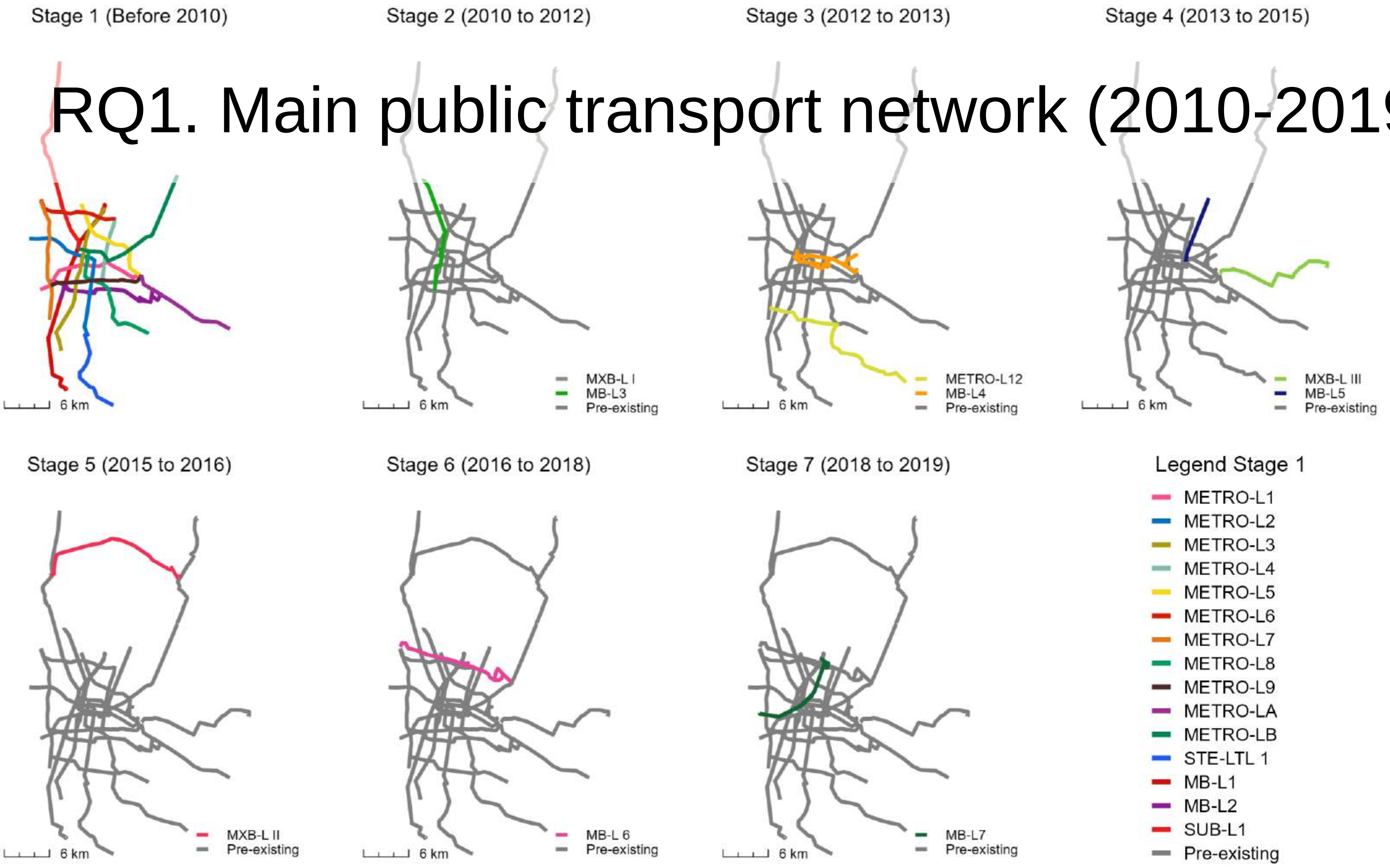
- Challenges:
  - Data-hungry: Model all-to-all travel time matrices, e.g. 1K x 1K = 1M
  - Until recently, only commercial software
  - For long time this has been the bottle-neck

Open-source software and big data

OpenStreetMap +
OpenTripPlanner +
PT timetables in ‘universal’ format (GTFS)
=
Detailed routes (time departure, transfers, variability of services)
Multiple scenarios (past, future, proposed or contrafactual)



RQ1. Main public transport network (2010-2019)

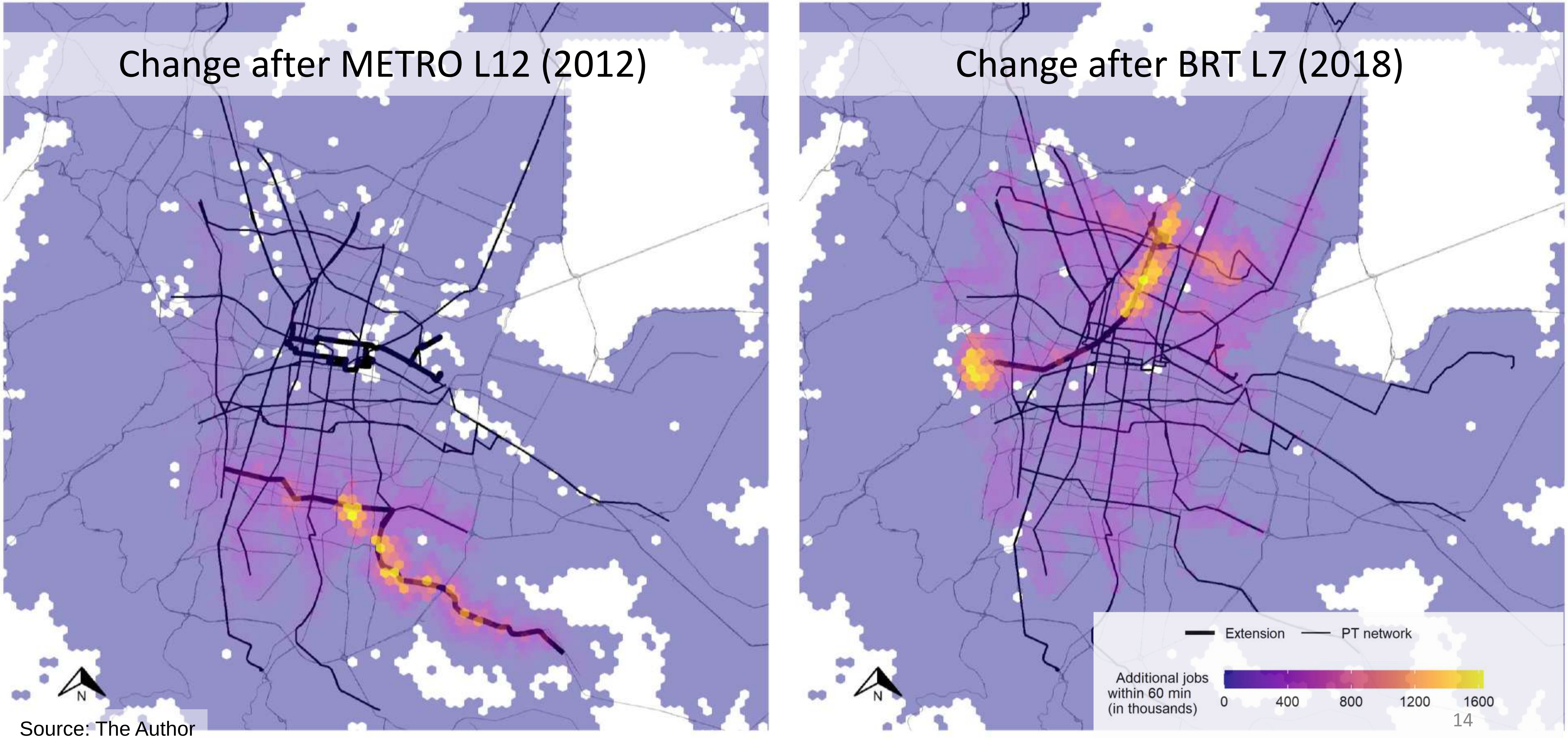


Source: The Author

RQ1. Accessibility – Travel time matrices

- Routes for:
  - Each origin to every possible destination (All-to-all)
  - 7 temporal scenarios
  - 8 times of departure: accounting for variability of the PT service
  - 5 zone schemes: accounting for spatial aggregations, e.g. MAUP
- ~1.3 billion routes by PT
- Plus ~230 million by car
- All free!
- Google?

RQ1. Accessibility change



Source: The Author

RQ2 - Methods

Research strategy:

1. Identify adequate accessibility measure and parameters
2. Estimate hedonic function considering spatial structure
3. Illustrate potential of LVC according to the introduction of a new BRT line (MB-L7)

Sample:

- Administrative mortgage records collected by *Federal Mortgage Society*
- Data used for national housing price index
- Size: N=~800K, from 2010 to 2019

All type of employment

$$A_{\mathrm{PT}jt} = \sum_{j=1}^{J} E_k^{\gamma} e^{-\beta_2 d_{jkt} + \mu \delta_{jk}}$$

Matching employment

$$A_{\mathrm{PT}jtm} = \sum_{j=1}^{J} E_{km}^{\gamma} e^{-\beta_2 d_{jkt} + \mu \delta_{jk}}$$

Parameters to be estimated

$$\{\gamma, \beta, \mu\}$$

RQ2 – Methods – Modelling framework

Non-linear model (estimate accessibility parameters)

$$\ln(P_i) = \alpha + \sum_{c=1}^{C} \beta_c \boldsymbol{x}_{ci} + \beta_{\text{APT}} \left(\sum_{j=1}^{J} E_k^{\gamma} \exp(-\beta_2 d_{jkt} + \mu \delta_{jk}) \right) + \epsilon_i$$

Spatial multilevel model – Besag-York-Mollie (BYM)

$$\ln(P_i) \sim N(\alpha + \beta_S^{\intercal} \mathbf{X}_{iS} + \beta_N^{\intercal} \mathbf{X}_{iN} + \beta_A^{\intercal} \mathbf{X}_{iA} + \beta_t Year_i + \beta_s State_i + \xi_{j[i]}, \sigma_e^2), \text{ for } i = 1, \ldots, n,$$

$$\xi_{j[i]} = \nu_j + \upsilon_j,$$

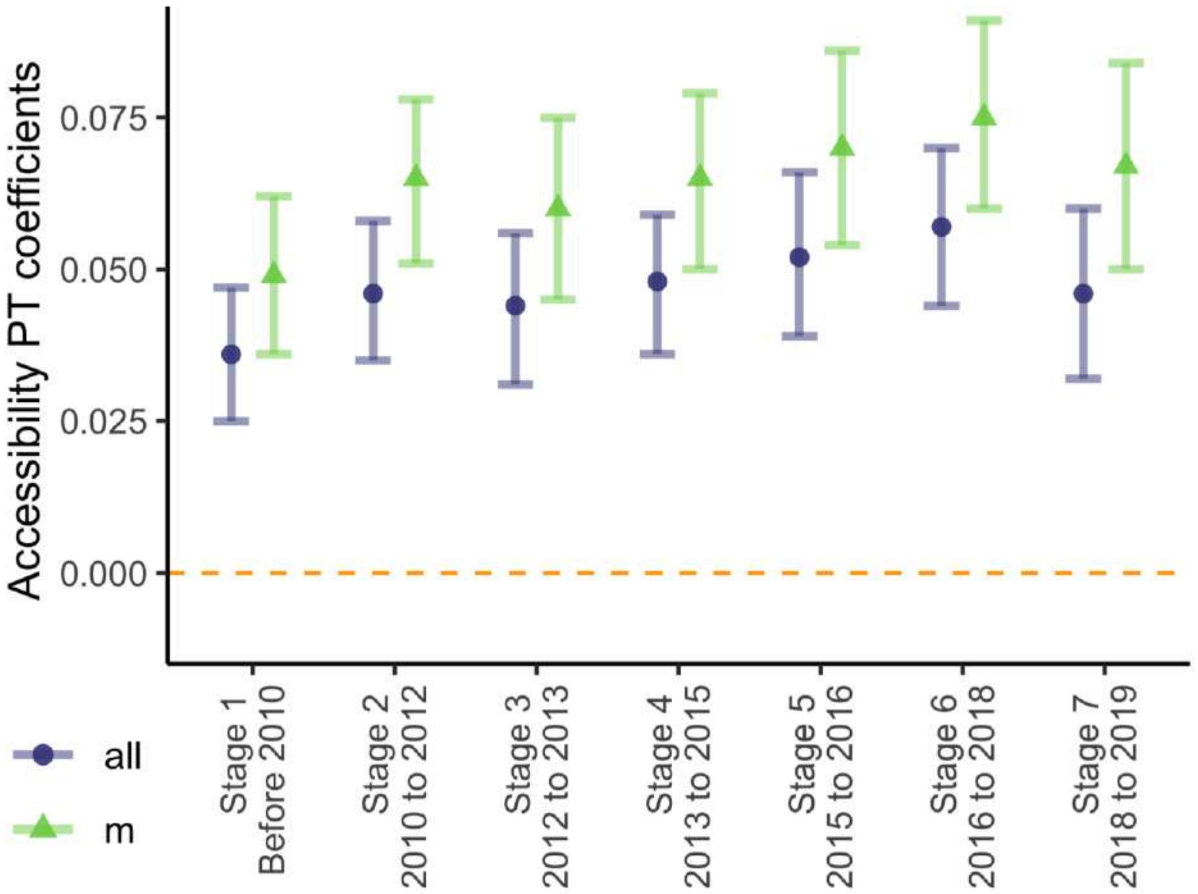
$$\nu_j \sim N(0, \sigma_\nu^2),$$

$$\upsilon_j \mid \boldsymbol{\upsilon}_{-j} \sim N\left(\frac{\sum_{j=1}^{J} w_{jk} \upsilon_k}{\sum_{j=1}^{J} w_{jk}}, \frac{\sigma_\upsilon^2}{\sum_{j=1}^{J} w_{jk}} \right),$$

$$\sigma_\nu^2, \sigma_\upsilon^2 \sim \text{Gamma}(a, b),$$

- Un-structured random effect
- Spatially structured random effect
- Intrinsic conditional auto-regressive (ICAR)

RQ2 – Results – Multilevel spatial model

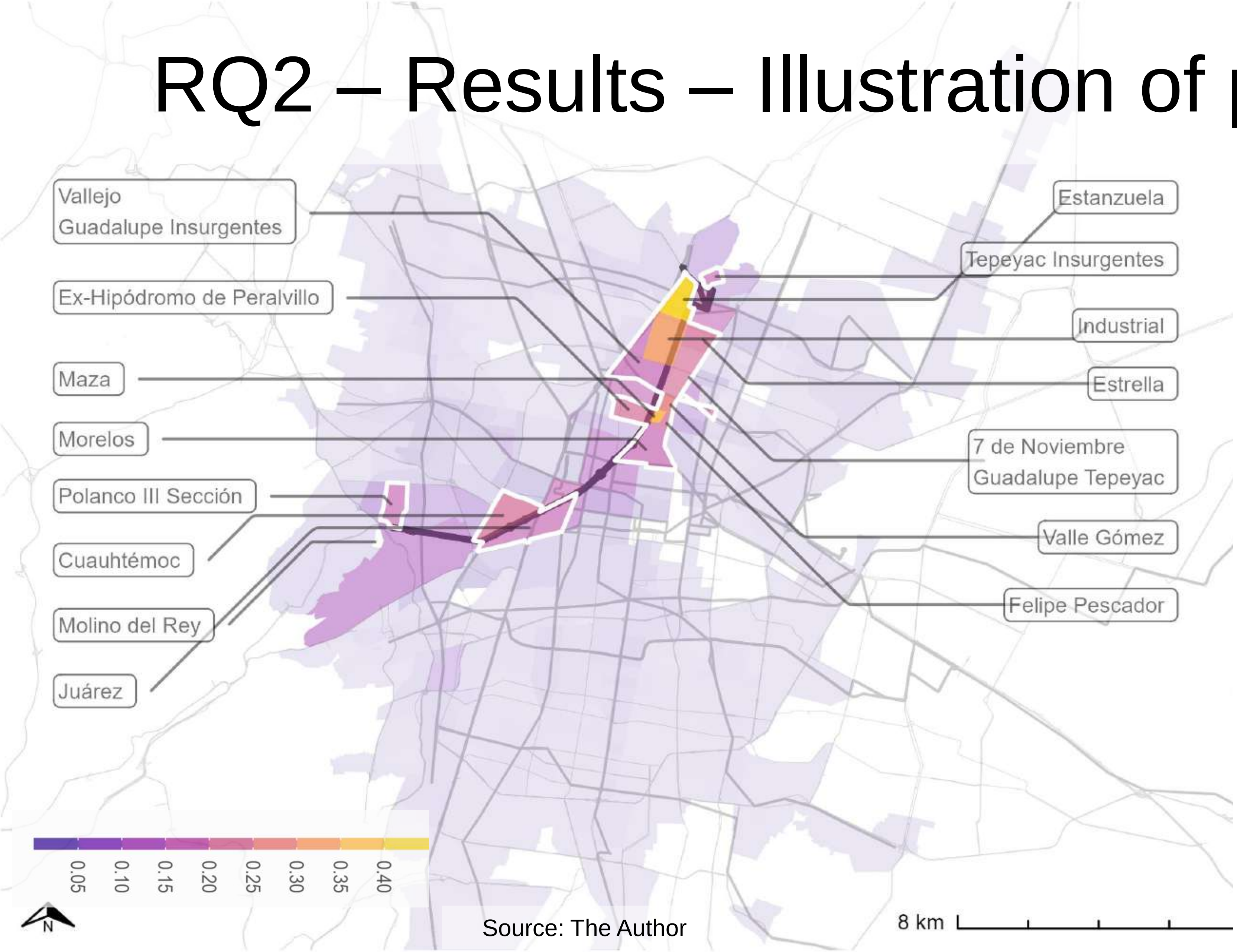


Coefficients for accessibility while controlling for structural and location characteristics (BYM)

Example:

- +¼ SD in accessibility (~13%) on a £75K dwelling:
  Increase of between £675 and £1,100 (depending on the Stage)

RQ2 – Results – Illustration of potential



Source: The Author

Aggregate benefits BRT-L7 (2018)

- Median housing value for each post code
- All affected housing stock: £157 M (1.5 times L7)
- Partial, only top 15 post codes: £69 M (0.7 times)

Conclusions

- Big data and open-source software are allowing us to re-address old questions and ideas.
- Consistent findings according to theory…These may contest previous neutral findings in the region.
- These resources are shown to be useful and relevant tools and are now available to quantitative researchers and policy analysts.
- Overall, encouraging results informing LVC-based polices

Thank you!

Do you have questions or comments?

Contact:

j.verduzco-torres.1@research.gla.ac.uk

Photo: Instagram @alan.kcs

References

1. (Giancarlo) Falcocchio, J. C., Malik, A., & Kontokosta, C. E. (2018). A data-driven methodology for equitable value-capture financing of public transit operations and maintenance. Transport Policy, 66, 107–115. https://doi.org/https://doi.org/10.1016/j.tranpol.2018.02.005
2. Murakami, J. (2012). Transit Value Capture: New Town Codevelopment Models and Land Market Updates in Tokyo and Hong Kong. In G. K. Ingram & Y.-H. Hong (Eds.), Value Capture and Land Policies (pp. 285–320). Cambridge, MA: Lincoln Institute of Land Policy. Retrieved from www.lincoln%5Cninst.edu/pubs/2026_Value-Capture-and-Land-Policies.
3. Transport for London. (2017). Land value capture: Final report. London. Retrieved from https://www.london.gov.uk/sites/default/files/land_value_capture_report_transport_for_london.pdf
4. McIntosh, J. (2015). Framework for land value capture from investments in transit in car-dependent cities. Journal of Transport and Land Use, 155–185. https://doi.org/10.5198/jtlu.2015.531
5. Liu, W., Wang, Q., & Wang, J. (2018). Research on the mechanism of value creation and capture process for urban rail development. Journal of Ambient Intelligence and Humanized Computing. https://doi.org/10.1007/s12652-018-1162-z
6. Salon, D., Sclar, E., & Barone, R. (2019). Can Location Value Capture Pay for Transit? Organizational Challenges of Transforming Theory into Practice. Urban Affairs Review, 55(3), 743–771. https://doi.org/10.1177/1078087417715523
7. Smolka, M. O., & Maleronka, C. (2018). Assessing the monetary relevance of land value capture: the case for charges for additional building rights in São Paulo , Brazil. International Journal of Real Estate and Land Planning, 1. Retrieved from https://ejournals.lib.auth.gr/reland
8. Smolka, M. O. (2013). Implementing Value Capture in Latin America: Policies and Tools for Urban Development.
9. Alonso, W. (1964). Location and Land Use: Toward a General Theory of Land Rent. Cambridge, MA: Harvard University Press. https://doi.org/10.1126/science.11.277.620
10. Higgins, C. D., & Kanaroglou, P. S. (2016). Forty years of modelling rapid transit's land value uplift in North America: Moving beyond the tip of the iceberg. *Transport Reviews*, *36*(5), 610–634. https://doi.org/10.1080/01441647.2016.1174748
11. Geurs, K. T., & van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: Review and research directions. Journal of Transport Geography, 12(2), 127–140. https://doi.org/10.1016/j.jtrangeo.2003.10.005

Appendix 1. LVC instruments

LVC Instrument	Rationale			Arrangement			Cost type		Contributor		Infrastructure		Land development	
	Direct	*Indirect*	*Macro*	*Compulsory*	*Negotiated*	*Voluntary*	*Capital*	*O&M*	*Landowner*	*Developer*	*Before*	*After*	*Existing*	*New*
Tax- or fee-based														
Land value tax or split rate tax	●	●		●			●	●	●		●	●	●	●
Betterment charges and SAD	●	●		●			●		●		●	●	●	
Tax increment financing		●		●	●	●	●		●		●		●	●
Development-based														
Land sales or leases			●			●	●	●	●	●	●	●		●
Air rights sales or leases	●	●				●	●	●		●		●		●
Joint development		●			●	●	●	●		●	●	●		●
Land readjustment or redevelopment schemes			●	●	●	●	●		●	●	●	●		●

Source: the Author based on Alterman, 2012; Suzuki et al, 2015; Zhao et al., 2012.

Appendix 2. Non-linear models

	M1:S2	M2:S2	M3:S2	M4:S2	M5:S2	M6:S2
	(OLS)	(OLS)	(NLM)	(NLM)	(NLM)	(NLM)
Accessibility PT	0.033***	0.045***	0.036***	0.044***	0.056***	0.066***
	(0.011)	(0.006)	(0.007)	(0.006)	(0.011)	(0.011)
$\hat{\beta}_2$			0.201***	0.053**	0.130***	0.099***
			(0.058)	(0.025)	(0.034)	(0.034)
$\hat{\mu}$			0.186	1.84*	3.54*	2.26**
			(1.09)	(1.01)	(1.86)	(1.01)
$\hat{\gamma}$					0.387***	0.337***
					(0.109)	(0.102)
Match opportunity	All	k	All	k	All	k
Structural controls	Yes	Yes	Yes	Yes	Yes	Yes
Locational controls	Yes	Yes	Yes	Yes	Yes	Yes
Dummy year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	781,898	781,898	781,898	781,898	781,898	781,898
Log-Likelihood	158,423.3	163,774.0	162,065.2	164,015.3	164,776.5	166,176.0
AIC	-316,776.7	-327,478.0	-324,056.4	-327,956.6	-329,477.0	-332,275.9
RMSE	0.1976	0.1962	0.1967	0.1962	0.1960	0.1956
RESET, statistic	13572.6	11635.4	12688.6	11700.2	14445.3	13501.3
VIF, mean	3.597	3.296	3.270	3.288	3.371	3.439
B-P test, statistic	62268.8	60962.5	63304.0	61230.0	61119.6	60196.9

Appendix 3. Full spatial BYM model results

	Stage 1 (Before 2010)		Stage 2 (2010 to 2012)		Stage 3 (2012 to 2013)		Stage 4 (2013 to 2015)		Stage 5 (2015 to 2016)		Stage 6 (2016 to 2018)		Stage 7 (2018 to 2019)	
Accessibility:	All	m	All	m	All	m	All	m	All	m	All	m	All	m
Accessibility PT	0.036***	0.049***	0.046***	0.065***	0.044***	0.060***	0.048***	0.065***	0.052***	0.070***	0.057***	0.075***	0.046***	0.067***
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.008)	(0.006)	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)	(0.009)
$\hat{\sigma}^2_y$	0.0184	0.0184	0.0182	0.0183	0.0186	0.0186	0.0213	0.0213	0.0200	0.0200	0.0224	0.0223	0.0231	0.0231
σ^2_v	0.0034	0.0028	0.0032	0.0041	0.0042	0.0041	0.0038	0.0040	0.0054	0.0035	0.0056	0.0056	0.0058	0.0066
$\hat{\sigma}^2_v$	0.0235	0.0241	0.0296	0.0262	0.0319	0.0312	0.0310	0.0303	0.0338	0.0382	0.0411	0.0371	0.0406	0.0387
Observations	146779	146779	116866	116866	79200	79200	118206	118206	73903	73903	142375	142375	104569	104569
Groups	1784	1784	1788	1788	1699	1699	1790	1790	1656	1656	1799	1799	1706	1706
Log-Likelihood	83239	83245	66184	66195	43701	43707	57736	57745	37973	37984	66474	66485	46744	46753
DIC	-168836	-168833	-134692	-134694	-89382	-89379	-117856	-117854	-77957	-77944	-135677	-135669	-95873	-95873
RMSE	0.1349	0.1349	0.1343	0.1343	0.1353	0.1353	0.1452	0.1452	0.1402	0.1402	0.1487	0.1487	0.1509	0.1509
Moran's I^a	0.7694***	0.7670***	0.7894***	0.7831***	0.7792***	0.7772***	0.8138***	0.8105***	0.8053***	0.8094***	0.8475***	0.8489***	0.8475***	0.8433***

Note:
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1; In Bayesian models 99%, 95% or 90% of the highest posterior density (HPD) credible interval does not include zero, respectevely. Posterior marginal mean reported with corresponding standard deviation in parentheses. All models include structural and year controls. DIC = Deviance information criterion. RMSE = Root mean squared error. Moran's I statistic on the upper-level random effect.

Appendix 4. Limitations

- Accessibility analyses do not consider *street transit* modes explicitly
- The property value data is constrained to the formal property market. Still useful since this data convers an important share of the total transactions (formal/informal) in the market (approx. 50%).
- Locational controls (e.g. employment, urban amenities) are estimated in one point of time due to the lack of more frequent update of data. These issue could affect OLS estimates but more severely than spatially structure models (BYM) since the latter are robust to omitted variables (Bivand et al., 2017; Lee, 2016)
- Statistical analyses are cross-sectional. Therefore, these do not allow causal inference.