Exploring distance decay pattern of public transport-induced agglomeration and its impacts on train ridership attraction

S Nurlaela$^{1,2}$, J (C) Xia$^{3,4}$, D Tuladhar$^{3,4}$, T (Grace) Lin$^{3,4}$ and P Lie$^{3,4}$

$^1$ Urban and Regional Planning Department, Faculty of Architecture Design and Planning
$^2$ Institut Teknologi Sepulung Nopember
$^3$ Department of Spatial Sciences, School of Earth and Planetary Sciences
$^4$ Curtin University of Technology, WA, Australia

Email: sitnurlael0@gmail.com; c.xia@curtin.edu.au; dinesh.tuladhar@postgrad.curtin.edu.au; zju.grace@gmail.com; peiyu.li@postgrad.curtin.edu.au

Abstract. Public transport infrastructure creates the effect of agglomeration through transportation externalities. Effective density is an accessibility based agglomeration that was raised as a positive externality from public transportation investments. The aim of this paper is to understand whether public transport facility would induce agglomeration around stations and furthermore induce train ridership. A methodology was developed to reveal the causality of effective density on ridership and reduce the confounding effects from land use-related determinant factor. This was shown by the propensity score matching that tested if effect of a station being in the treatment group (high effective density stations) on train ridership was influenced by land use characteristics of catchment stations. The causality of effective density on ridership was compared between station groups. Findings showed the effect of treatment group was higher in the matched sample compare to the unmatched sample. This difference may be assigned as the true effect of public transport induced agglomeration which was higher after controlling the land use characteristics of stations. Thus, the inclusion of land use variables in the model prediction may has the effect of rendering the influence of effective density variable lower in the model. These findings could guide station catchment area planning to maximise effective density benefits on train ridership.

Keywords: effective density, propensity score matching, agglomeration, public transport infrastructure, train ridership.

1. Introduction

The impacts of transportation systems in terms of transportation infrastructure need to be evaluated not only on their aspect of land use but also travel behavior [1]. Further, other factors such as the spatial economic dimension may come into play when transportation system-impacted land use and travel behavior to be examined in a holistic manner.

The extension of the new Perth – Mandurah railway line in the Perth Metropolitan Region (PMR) has been assumed to have made a direct influence on land use and property development, or on economic development. The question of the contribution of stations along the new line to increased train ridership has been considered for more than ten years. Public transport infrastructure creates the effect of agglomeration through transportation externalities. Public transport-induced agglomeration
economies, which can also be considered to be part of technological externalities, may be defined as the concentration of economic activities and the clustering of offices, shops, entertainment centers, and other land uses that emerge around public transportation stops. The benefits from this clustering are increased efficiency through lower infrastructure costs and reduced labor costs, while more opportunities are created for greater access to specialized labor [2].

Effective density is a specific concept of agglomeration that were raised as a positive externality from public transportation investment ([3], [4]). Transport investments may induce positive productivity benefits by forming clusters of economic activities which are more accessible (effective density) [5]. The improved transport may lower travel time or costs, reduce production cost and help growth a region via effective density. However, it is not clear how to explain the spatial relationship between transport provision and the scale of effective density. In other words, what is the spatial distribution of effective density around transport facilitates, such as train stations? Is closer to a train station, more economic activities or clusters of firms or employees occurred, thus, would be more accessible by train, as such intended by the Transit Oriented Development concept? And what does it imply on travel behavior such as transit demand?

This paper aims to understand whether public transport facility would induce agglomeration around stations and furthermore induce train ridership. This paper explores the spatial pattern or characteristics of effective density attached on public transport induced agglomeration. As agglomeration emerged majorly near a transport facility, i.e. a train station, one may expect its spatial distribution to form a distance decay pattern from a station. Distance decay of the effective density may reflect the interaction of transport supply, such as a train station and land use in the surrounding areas, such as stations’ catchment areas; whereas effective density attenuate with distance away from a station [6].

2. Agglomeration measurement
The public transport induced agglomeration concept was initiated by [3] and further developed by ([4]). The study of [3] highlighted the relationship between dense spatial units and the clustering of economic activity with an increase in productivity. Venables’ original concept explained the relationship between wages, travel costs, and land rent or housing costs and its impacts on city size. There is a trade-off between land rent or housing costs and commuting costs as: “workers located closer to the CBD face lower commuting costs but higher rents…” [3, p.9]. Venables added an endogenous productivity effect in his model. In Venables’ model formulation, transport improvement (such as the development of new railway line and stations) was modelled as a reduction in commuting costs, which derived a further effect on the real income. As workers were able to save more money (a conversion from travel costs savings), workers could afford to live closer to the centre. Transportation externalities therefore create increased urban productivity, which in turn increase the city size since “the city expands up to the point at which these are high enough that a worker is indifferent between locating at the edge of the city and commuting to the CBD, or living (and working) in a non-city location” [3, pp. 3-4].

Graham, after Venables, modelled agglomeration economies using a measure that incorporates both proximity (accessibility) and the scale of economic activity (the size of employment), what so called as the effective density [4]. The total effective density of employment defined in Graham is a measurement of agglomeration. Specifically, agglomeration in terms of effective density was measured in two ways: by using Euclidean distance or network distance [7]. This paper applied the measurement of effective density based on a network travel time.

The formula for effective density calculation was specified as follows [4, p. 327]:

\[
ED_{lm} = \frac{E_i}{\sqrt{A_i/\pi}} + \sum_{j \neq i} \frac{E_j}{td_{ij}^{\text{eff}}} 
\]
Where:

- \( ED_{im} \) = the employment density of ward \( i \) for industry in sector \( m \).
- \( E_i \) = the number of employment of ward \( i \)
- \( A_i \) = the land area of ward \( i \)
- \( E_j \) = the number of employment of ward \( j \)
- \( t_{dij}^\alpha \) = the travel distance between ward \( i \) and ward \( j \), weighted by the distance decay parameter \( \alpha \).

3. Methodology

Although there is a strong theoretical basis for explaining how better transport infrastructure helps achieve the scale of economies of agglomeration, limited research has been conducted to understand the distance decay patterns of agglomeration around train stations and the factors affecting such patterns [8]. On the other hand, the study of train ridership prediction had not yet paid much attention on the spatial-economic factor, such as effective density or worker/job supply. Nonetheless, how both agglomeration of effective density (the scale or concentration), the proximity, and its decay pattern from stations could increase train ridership was not known in the literature.

In this research, stations were investigated for their effective density pattern, the level of land use development, and train ridership magnitude. A station classification was derived based on these three factors, dividing stations into a control group and a treatment group where the level of land use development were controlled or matched between these two groups.

The model framework is shown in Figure 1. It describes the relationships between a train station (a node) and its catchment area (a place) following a node-place model [9]. To capture a distance decay effect, a larger catchment is necessary as a unit of analysis. The park and rider catchment area was used to test the research hypothesis on public transport induced agglomeration. The black dotted line in Figure 1 illustrates the decrease of effective density with distance away from a train station located at the centre of the diagram.

Two factors were defined in spatial distance decay interaction according to the predecessors’ research [10]. These factors would define the determinant of train ridership.

- Interaction behaviour factors, represented characteristics in the relationship between suburbs in the catchment area and between a catchment area of a station (park and ride catchment) and its corresponding station. Variables consisted of the decay parameter of effective density, the level of public transport supply index for suburbs in the catchment area, road network travel distance or a centrality of a suburb to all other suburbs that was measured by road network, distance of suburbs in catchment area on average to the corresponding station.
- Spatial structure: refers to the size and configuration of spatial system. It refers to the number of job, land rent, job-housing balance, wage level on average of each suburb included in a catchment area.

To test the extent to which the concentration and/or decay pattern of effective density influence train trip attracted to each station, stations were classified into a quadrant matrix (figure 2). This research assumed stations in quadrant 1 or q1 group (negative decay parameter and effective density above mean of Perth Metropolitan region) would have higher train trip attraction than other stations. Less negative decay parameter means less travel friction or higher accessibility between station and its catchment, thus, larger size of catchment station and higher employed resident opportunity. However, a positive decay parameter reflect no or less concentration in the station area, or area other than stations had become a more developed area than that of the station area itself.
4. Method

4.1. Data collection
Most of the data used in this research were collected from various government agencies and Australian Bureau of Statistics (ABS):
• This research used ABS journey to work data as dependent variable in main analysis of train ridership model.
• Transportation information, such as travel time between suburbs, was derived from the 2011 Strategic Transport Evaluation Model (STEM) by the Department of Planning of Western Australia.
Australia, consisted of 472 times 472 travel zones in the system in which these travel zones were converted into suburb matrices, consisted of 317 x 317 suburbs. The travel time by park and ride journey was chosen as the best alternative to represent the spatial interaction of the train trips and to sufficiently captured the indication of distance decay pattern.

- Job data was retrieved from the ABS website, sourced from 990 DZN or destination zones in the Perth Metropolitan region that also were converted into suburb unit. Employed resident data was sourced from the ABS censuses based on suburb database.

4.2. Study area
This research used the suburb or the state suburb (SSC) as an administrative boundary of the study area. The SSC was not part of the ABS structure. Instead, the areas were aligned closely with the Statistical Areas level 2 (SA2), where the SA2 are an aggregation of statistical area level 1 (SA1). Both SA1 and SA2 are defined in the ABS structure (2011 Census Dictionary, p. 183 and p. 189).

Perth Metropolitan Area (PMA) has five train lines: Fremantle, Midland, Armadale, Mandurah and Joondalup lines and 70 train stations. Study area consisted of the stations as nodes and the catchment area in terms of suburbs as places.

![Figure 3. The map of study area](image_url)

4.3. Data analysis
To form a station classification, an exponential distance decay model was performed. According to figure 1, the exponential distance decay model depicted the relationship between the magnitude of effective density and the distance of the catchment area to the station. A set of significant decay parameters was derived. Only 45 stations had significant parameters. These 45 stations were allocated into the quadrant according to its decay parameter value and its effective density value.

Propensity score matching (PSM) was used to investigate whether the research hypothesis that a station being in the treatment group or q1 group (has distance decay pattern of agglomeration with high value of effective density) was a predictor of train ridership, when controlling for land use variables. These land use data were generated at the catchment station level by using data manipulation in GIS. There were 68 stations being research in the study area from total 70, two were removed due to
outliers in the employment data. The distance decay model resulted in 45 stations having a significant parameter of exponential distance decay, thus, 45 stations being sampled in the PSM. Two stations with the outlier distance decay parameter were removed, resulting in 43 stations for the final analysis. The quadrant matrix divided these 43 stations into treatment and control group according to the criteria of value of distance decay parameter (negative or positive) and the value of employed resident effective density (above mean or below mean). The quadrant matrix derived 15 stations in the q1 group and 28 stations in other groups for train trip attraction model. As the matching variables, this research choose land use variables such as the number of job and land (property) value. The rationale for choosing those matching variables was that the number of job and land rent may increase due to both the urbanization (population development) and the improved accessibility. Land use development could be indicator of indirect productivity, a proxy for agglomeration effect and urbanization. The area with more intensive land use development was assumed to have higher activities, denser, more expensive land or property prices, and higher wage offered, thus, had higher agglomeration and/or urbanization. Controlling these two variables resulted in assigning the public transport induced agglomeration as only a function of proximity or improved accessibility post the railway line extension, therefore it being modelled by adding the q1 group as a dummy variable in the regression of train ridership model. By matching on those variables, the goal was to reduce the confounding influence of land use variables on the effect of public transport induced agglomeration on train ridership. As two regressions for train ridership model were imposed (before and after PSM), how much the regression model improved after the PSM could be assessed. The improved regression model after the PSM may reflect the true influence of public transport induced agglomeration on train ridership.

5. Results and discussion

5.1. Exponential decay model and station classification

Exponential distance decay model was performed to classify station based on hypothetical relationship between effective density and train ridership. To perform this model, first, the relationship between station and its catchment area was constructed in a fishnet structure to derive many observations than otherwise if only suburbs were used as observations. Secondly, an exponential distance decay model was performed between the effective density and the distance of catchment area to the station for each fishnet unit. An example of the exponential distance decay model from 5 stations was displayed in table 1 (completed data consisted of overall 45 stations with significant value of decay parameters).

| Station             | Beta*1000 | Beta_edertt | R sq | P value | Constant | Number of fishnets |
|---------------------|-----------|-------------|------|---------|----------|-------------------|
| Armadale Stn        | 2         | 0.002       | 0.398| 0       | 7.806    | 158               |
| Cannington Stn      | 1         | 0.001       | 0.365| 0       | 12.986   | 60                |
| East Guildford Stn  | -0.54     | -0.00054    | 0.297| 0       | 9.988    | 541               |
| Grant Street Stn    | 9         | 0.009       | 0.742| 0       | 10.42    | 27                |
| Stirling Stn        | -2        | -0.002      | 0.464| 0       | 16.613   | 151               |

Overall, the negative decay parameter ranged from -0.00055 to < 0 and the positive parameter ranged from > 0 to +0.004. Claisebrook station and Grant street station were excluded from further analysis due to outlier in the beta parameter, i.e. 0.015 and 0.09 respectively. This has resulted in 43 stations in total being used in the PSM model. Negative decay parameter means the number of effective employed density reduced for every one kilometre increase in the distance of catchment to the station. For example, for the catchment area in Stirling station, an area 1 km closer to the station had 0.2% more effective employed resident density. On the other hand, an area 1 km closer to the
Warnbro station had 0.1% less effective employed resident density, means the core of development was not emerged in the Warnbro station catchment, but on other area farther away.

The result of the exponential model being combined with the information on the magnitude of effective employed resident density constructed a station classification (figure 2). The number of station ridership in terms of the proportion of train trip attraction was added into the graph in the form of bubble size. It was worth noting that some of the stations with the highest train trip attraction did locate in the quadrant 1 (the treatment group), indicating the research hypothesis that a chance for a station being in the q1 group may have higher influence on train trip attraction, may be in evidence. The result is shown in figure 4. To understand the true impact of public transport induced agglomeration on train ridership, i.e. indicated by the probability of a station being in the q1 group, there was a need to control for the level of development in each station, such as represented by the number of job and land rent. At this stage, PSM analysis was used.

![Figure 4. Station classification in a quadrant matrix based on three variables: effective employed density (the y axis), distance decay parameter (the x axis), and the train trip attraction (the bubble size).](image)

### 5.2. Propensity score matching

The question of interest was whether the station classified in the group of negative distance decay – effective density above mean of Perth (group 1 or treatment group) compared to other station categories (control group) as displayed in the quadrant matrix in fig.4 has any higher impact on the proportion of train trip attraction assessed post the Perth-Mandurah railway line extension. Observed differences in the proportion of train ridership attraction may be due to land use characteristics that has been presented before the extension. If the land use differences had been existed prior the railway extension, then the estimation of the true causal effect of effective density on train ridership will be biased due to confounding. That is, the public transport induced agglomeration as assumed at the first place, may not be hold. Thus, controlling for the similar level of land use development between the treatment group (group 1) and the control group (group 2) means assigning the influence of public transport induced agglomeration on train ridership as a function of improved accessibility post the railway extension that lowering commuting costs.
A propensity score matching analysis was used to control for these confounding influences. The tasks were conducted in the SPSS software with extension packages. In a first step the propensity score, i.e. the probability of a station to fall into the group 1 was estimated using logistic regression. All covariates from land use variables were used. After the estimation of the propensity score, the stations were matched using a 1 to 3 nearest neighbor matching. In order to exclude bad matches, a caliper of 0.2 was imposed. After matching, the balance of all observed covariates was examined. There was no covariate exhibits a large imbalance at \( d>0.25 \). The covariate balance was improved in the matched sample. The actual propensity score distributions of both groups (treatment and control) before and after matching overlaid with a kernel density estimate. The \( t \)-test analyses were carried out to test the differences in the standardized mean value of variable of job number and land rent out of the treatment and control group. It showed that after the PSM, the adjusted estimate of the mean differences in job number and land rent between the two samples were insignificant statistically. This means the PSM has successfully balance the covariates of those two variables in the two samples. Thus, stations in the treatment and control group have similar land use development or land use characteristics according to the number of job and land rent.

5.3. Train trip attraction model
Regression model before and after the PSM were discussed to understand if controlling land use variables could inform the true impact of public transport induced agglomeration, i.e. the commuting cost reduction due to travel time improvement (post the Perth-Mandurah railway extension) in the system may influence decision of more workers to use train as mode of journey to work travel.

The comparison of the model was assessed based on several \( t \)-tests. These comprised of the assessment of the mean proportion of train trip attraction before and after matching and the effect of dummy variable \( q1 \) group in the model (group 1: a station being in a group of quadrant 1: negative distance decay and high effective density). The model improvement was compared based on the adjusted R-square.

The model before and after the propensity score matching was compared (table 2 and 3). The variables of interests were the dummy variable \( q1 \) group, the distance of catchment from station, the magnitude of effective density, and the interaction variable between effective employed resident density and the station distance.

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|---|----------|-------------------|--------------------------|
| 1     | .774* | .599     | .489              | .04682                   |

a. Predictors: (Constant), dtedertt, wage_b, job3, q1 group, stri, lvr, ederrt3, pti, dist
b. Dependent Variable: ptrainw

| Coefficients* | Model | Unstandardized | Standardized |
|---------------|-------|----------------|--------------|
|               | B     | Std. Error    | Beta         |
| 1 (Constant)  | .437  | .179          |              |
| \( q1 \) group (dummy variable of treatment group) | .053  | .024          | .391         | 2.187 | .036** |
| wage_b (hourly wage level) | -.011 | .006          | -.276        | -1.707 | .097* |
| Lvr (land rent per sq meter) | 4.252E-5 | .000         | .541         | 2.316 | .027** |
| Job (the number of job) | 1.032E-5 | .000         | .204         | 1.084 | .286 |
| Dist (distance to the nearest station) | .027  | .019          | 1.112        | 1.413 | .167 |
| Stri (road network accessibility) | -8.861E-6 | .000       | -.428        | -1.263 | .216 |
| Pti (public transport accessibility) | -2.613E-8 | .000       | -.322        | -1.055 | .299 |
| Ederrt (employed resident effective density) | -.012 | .009          | -.305        | -1.266 | .214 |
| Diederrt (the interaction between distance and effective density) | -.005 | .002          | -.968        | -1.921 | .063* |

a. Dependent Variable: ptrainw

* ***) Significant at 99% level of confidence
   **) Significant at 95% of confidence
   *) Significant at 90% level of confidence
### Table 3. Train trip attraction model after the PSM

| Model Summary* | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|----------------|---|----------|-------------------|--------------------------|
| 1              | .814 | .663 | .529 | .04799 |

Model Summary

- a. Predictors: (Constant), dtedertt, q1group, wage, b, Propensity Score, stri, pti, edertt3, dist
- b. Dependent Variable: ptrainw

| Coefficients* | Unstandardized Coefficients | Standardized Coefficients | t | Sig. |
|---------------|-----------------------------|---------------------------|---|-----|
| Model         | B                           | Std. Error                | Beta |      |     |
| 1             |                             |                           |     |     |
| (Constant)    | .562                        | .227                      | 2.479 | .022**|
| q1group       | .072                        | .023                      | .501 | 3.079 | .006**|
| Propensity Score | .095                      | .057                      | .343 | 1.661 | .112 |
| wage, b       | -.009                       | .008                      | -.207 | -1.092 | .288 |
| Dist          | .043                        | .028                      | 1.460 | 1.544 | .138 |
| Stri          | -1.268E-5                   | .000                      | -5.85 | -1.489 | .152 |
| Pti           | -3.189E-8                   | .000                      | -3.17 | -.923 | .367 |
| edertt3       | -.009                       | .013                      | -.226 | -7.20 | .480 |
| Dtedertt      | -.006                       | .002                      | -1.380 | -2.545 | .019**|

 Dependent Variable: ptrainw

* *** Significant at 99% level of confidence
** ** Significant at 95% of confidence
*) Significant at 90% level of confidence

The effect of public transport induced agglomeration in the regression model was represented by dummy variable q1group. T-test analysis conducted before the PSM showed the unadjusted estimate of the effect of public transport induced agglomeration as measured by the q1group was statistically significant (t(43)=2.187, p<0.05). Increasing one standard score on the chance of a station of being in the q1 group would increase the proportion of train trip attracted to this station by 0.391 of its standard score. Stations that classified in the q1 group attracted higher proportion of train trip attraction (ptrainw=13.3%) than stations in other groups (ptrainw=6.04%) and the standardized mean difference was d=7.25%. This estimated effect was not necessarily representing the true causal effect of public transport induced agglomeration. There may be many covariates act as potential confounders that can bias the estimated effect of effective density on train ridership.

Further, the regression model after the PSM was conducted, now included a total of 21 stations in the matched sample. The stations were almost evenly distributed in the two groups (10 stations for the treated and 11 stations for the control) or 29 stations according to the paired sample with ratio 1:3 matching scheme. The adjusted estimate of the effect of q1group on train trip attraction in the matching sample was still significant (t = 3.079, p=0.006 or <0.01). Increasing the chance of a station for being in the q1 group by one standard score, would increase the proportion of train ridership attraction to this station by 0.501 of its standard score. The t-test in the differences on the proportion of train trip attraction for stations in the matched sample calculated for group 1 and group 2. Finding showed that stations in the matched sample classified in the q1 group had significantly higher proportion of train trip attraction (ptrainw=12.39%) than that of stations in other groups (ptrainw=6.8%) at p=0.019. The standardized mean difference in the two samples was 5.5%.

There has been an improvement in terms of R-square increases, or of prediction capability of approximately 10.68% relatively (from 59.9% to 66.3%) in the regression after PSM. The adjusted R-square in the model post PSM indicated 52.9% of the variation in the proportion of train trip attraction may be explained by the model, when incorporating the adjustments of the number of independent variables and the sample sizes. Both the dummy variable q1 group and the interaction variable between effective density and distance to station in the model after the PSM were significant. These two variables informed the effect of public transport induced agglomeration on train ridership and how much this effect changed as a decay from a station.

Increasing the chance of a station of being in the q1 group (has distance decay pattern and high effective density) by one standard deviation increased the proportion of train trip attraction by 0.501 of
its standard score. The effect of q1group variable was higher in the matched sample compare to the unmatched sample. The effect of q1group in the unmatched sample (43 stations) was 0.391. The effect of 0.501 may be assigned as the true effect of public transport induced agglomeration which was higher after controlling the land use characteristics of stations. More details on the modeling results can be requested from the authors.

6. Conclusion

This paper was expected to shed light on any important interrelationship between transportation, land use, and agglomeration in terms of effective density for the case of train ridership modelling. Understanding to what extent effective density influence train ridership and that this impact is differentiated from the influence of land use related determinant factor of ridership would give different implications for transportation policies. These results inform the significant and important aspect of accessibility based agglomeration measurement in the policy analysis to optimizing the benefits of public transport externalities to residents, businesses and communities.

7. References

[1] Brons, M., M. Givoni, et al. (2009). "Access to railway stations and its potential in increasing rail use." Transportation Research Part A: Policy and Practice 43(2): 136-149.
[2] Weisbrod, G. & A. Reno. 2009. Economic impact of public transportation investment. In Preparation report for TCRP Project J-11, Task 7, Transit Cooperative Research Program. the American Public Transportation Association.
[3] Venables, A.J (2004). Evaluating Urban Transport Improvements: Cost-Benefit Analysis in the Presence of Agglomeration and Income Taxation. CEF Discussion Paper No. 651. Center for Economic Performance.
[4] Graham, D. J. (2007) “Agglomeration, Productivity and Transport Investment”. Journal of Transport Economics and Policy, 41, 317-343.
[5] Venables, A. J. (2007) “Evaluating urban transport improvements: cost-benefit analysis in the presence of agglomeration and income taxation”. Journal of transport economics and policy, 173-188.
[6] Hansen, W. G. (1959) How accessibility shapes land use. Journal of the American Institute of Planners, 25, 73-76.
[7] Graham, D. J. & P. C. Melo. (2010). Advice on the Assessment of Wider Economic Impacts: a report for HS2.
[8] Graham, D. J., S. Gibbons & R. Martin (2009) Transport investments and the distance decay of agglomeration benefits.
[9] Bertolini, L (1999). “Spatial development patterns and public transport: the application of an analytical model in the Netherlands”. Planning Practice and Research, 14 (2), 199 – 210.
[10] Fotheringham, A. S. (1981) Spatial Structure and Distance-Decay Parameters. Annals of the Association of American Geographers, 71, 425-436.

Acknowledgements

Authors are grateful to the Department of Planning (WA) for providing travel time matrices data from the 2011 STEM model and the Landgate (© Western Australian Land Information Authority) for providing 2006 – 2011 property value transaction database for Perth metropolitan region.